ies NATIONAL CENTER FOR
EDUCATION STATISTICS
Institute of Education Sciences

U.S. Department of Education
NCES 2007–349

Documentation for the 2004–05 Teacher Follow-up Survey

March 2007

ies NATIONAL CENTER FOR
EDUCATION STATISTICS
Institute of Education Sciences

ies NATIONAL CENTER FOR
EDUCATION STATISTICS
Institute of Education Sciences

SASS

Schools and Staffing Survey

U.S. Department of Education
NCES 2007–349

Documentation for the 2004–05 Teacher Follow-up Survey

March 2007

Shawna Cox
Randall Parmer
Steven Tourkin
Toni Warner
U.S. Bureau of the Census

Deanna M. Lyter
Education Statistics Services Institute— American Institutes for Research

Reneé Rowland
Project Officer
National Center for Education Statistics

U.S. Department of Education
Margaret Spellings
Secretary

Institute of Education Sciences
Grover J. Whitehurst
Director

National Center for Education Statistics
Mark Schneider
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to

> National Center for Education Statistics
> Institute of Education Sciences
> U.S. Department of Education
> 1990 K Street NW
> Washington, DC 20006-5651

March 2007

The NCES World Wide Web Home Page is http://nces.ed.gov.
The NCES World Wide Web Electronic Catalog is http://nces.ed.gov/pubsearch.

Suggested Citation

Cox, S., Parmer, T., Tourkin, S., Warner, T., and Lyter, D.M. (2007). *Documentation for the 2004–05 Teacher Follow-up Survey* (NCES 2007–349). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

For ordering information on this report, write to

> U.S. Department of Education
> EDPubs
> P.O. Box 1398
> Jessup, MD 20794-1398

Or call toll free 1-877-4ED–Pubs or order online at http://www.edpubs.org.

Content Contact:
Reneé Rowland
(202) 502-7336
renee.rowland@ed.gov

Contents

List of Tables

List of Exhibits

Chapter 1. Overview

The Teacher Follow-up Survey (TFS) is sponsored by the National Center for Education Statistics (NCES) on behalf of the U.S. Department of Education and is conducted by the U.S. Census Bureau.

TFS is a follow-up survey of selected elementary and secondary school teachers who participated in the Schools and Staffing Survey (SASS).[1] SASS is the largest, most extensive survey of kindergarten through 12th grade (K–12) school districts, schools, teachers, and administrators in the United States today. It provides data on the characteristics and qualifications of teachers and principals, teacher hiring practices, professional development, class size, and other conditions in schools across the nation. TFS focuses on a sample of teachers who participated in SASS, including those teachers who left the K–12 teaching profession and those who continue to teach.

TFS includes teacher data from public (including public charter) and private schools, similar to SASS. However, TFS does not include teachers who taught in a school funded by the Bureau of Indian Affairs (BIA) during the SASS school year due to insufficient sample sizes in TFS. Together, SASS and TFS data provide a multitude of opportunities for analysis and reporting on elementary and secondary educational issues.

Background

TFS is a follow-up of selected teachers from the SASS teacher surveys and is conducted during the school year following the SASS administration. It was conducted in the 1988–89, 1991–92, 1994–95, 2000–01, and 2004–05 school years (after the 1987–88, 1990–91, 1993–94, 1999–2000, and 2003–04 administrations of SASS, respectively). NCES currently plans to conduct the next survey in the 2008–09 school year; it will collect data from a subsample of teachers who participate in the 2007–08 SASS.

Over time, the philosophy behind TFS has changed. The 2004–05 TFS more closely resembled the SASS teacher questionnaires than in any of the previous TFS administrations. There was a greater overlap of TFS and SASS teacher items, and there were fewer items unique to TFS, other than items pertaining to leaving last year's teaching position. When examined together, the results of TFS and SASS can give researchers insight on many different educational issues, including the retention of teachers in public and private schools and teachers' job satisfaction.

Congress, state education departments, federal agencies, private school associations, teacher associations, and educational organizations have used data from the 1988–89, 1991–92, 1994–95, and 2000–01 surveys. In particular, results of these prior administrations have been used to analyze changes in the teacher labor force over time, to develop incentive programs to encourage teacher retention, and to understand the effects of school practices and policies on a teacher's decision to continue teaching or leave the K–12 teaching profession.

Purpose and Content of the Survey

TFS is a one-time follow-up to the SASS teacher questionnaires and is conducted during the school year following SASS. The major objectives of the 2004–05 TFS are to measure the attrition rate for teachers, examine the characteristics of teachers who stayed in the teaching profession and those who left or retired, obtain activity or occupational data for those who left the position of a K–12 teacher, obtain

---

[1] For a complete description of the 2003–04 Schools and Staffing Survey, see *Documentation for the 2003–04 Schools and Staffing Survey* (Tourkin et al. 2007).

current teaching assignment information for those who are still teaching, and collect data on attitudes about the teaching profession in general and job satisfaction in particular. All basic demographic information about each teacher (e.g., race and ethnicity information) is collected on the SASS teacher questionnaires. TFS data can be linked to SASS data to provide contextual data on relationships between local districts and school policies and practices, teacher characteristics, and teacher attrition and retention.

TFS is composed of two questionnaires: the Questionnaire for Former Teachers (Form TFS-2), which collects information about sampled teachers who left the K–12 teaching profession after the 2003–04 school year, and the Questionnaire for Current Teachers (Form TFS-3), which collects information on sampled teachers who currently teach students in any of grades prekindergarten through 12.

Questionnaire for Former Teachers (Form TFS-2)

The purpose of the 2004–05 former teacher questionnaire was to obtain information about those respondents who left teaching within the year after SASS, such as their present occupation or activity, educational plans, reasons for leaving teaching, intent to return to teaching, attitudes about teaching, and demographic characteristics.

The 2004–05 Questionnaire for Former Teachers had the following six sections:

- *Section I—Employment Status* collected general information about employment, salary and supplemental income, attitude on remaining in current position, retirement, and retirement plans.
- *Section II—Information on Leaving the Teaching Profession* obtained information about the factors that influenced the decision to leave the position of a K–12 teacher.
- *Section III—Your Impressions of Teaching and of Your Current Job* collected information about the performance of last year's principal or school head, state or district assessment programs, and satisfaction with current job relative to teaching.
- *Section IV—Education Activities and Future Plans* collected information about recent enrollment in college or university courses and the possibility of returning to the position of a K–12 teacher in the future.
- *Section V—Background Information* obtained information about family income, household characteristics, marital status, and internet access.
- *Section VI—Contact Information* requested that respondents provide their personal contact information as well as contact information for two additional people who would be able to reach them in the event that they relocate.

Questionnaire for Current Teachers (Form TFS-3)

The purpose of the 2004–05 current teacher questionnaire was to obtain information about current teachers, including teachers who continued to teach in the same school as in the previous year and those who changed schools. It contains information about their current teaching assignments, attitudes about teaching, attitudes about last year's school, expected duration in teaching, plans for further education, and demographic characteristics.

The 2004–05 Questionnaire for Current Teachers had the following six sections:

- *Section I—Certification and Assignments at Your Current School* collected information about general teaching status, subjects taught, teaching certification, grade levels taught, organization of classes, students taught having an Individual Education Plan (IEP), and students taught who had limited-English proficiency (LEP).

- *Section II—Your Current School: Conditions and Experiences* obtained information on experiences and working conditions at the current school, student problems, school problems, attitudes and perceptions about teaching at the current school, hours worked and time allocation, and participation in additional school-related activities.
- *Section III—Information About Changes From Last School Year to This School Year* collected information about whether the teacher is teaching at the same school as the previous year, general information about the new school (if the teacher changed schools), factors that influenced the decision to leave the previous school (if the teacher changed schools), performance of last year's principal or school head, attitudes and perceptions about state or district assessment programs at last year's school, and satisfaction with current teaching position relative to last year's teaching position.
- *Section IV—Education Activities and Future Plans* collected information about recent enrollment in college or university courses, attitude about remaining in teaching, attempt to leave the position of a pre-K–12 teacher, and retirement.
- *Section V—General Employment and Background Information* obtained information about teacher salary and supplemental income, family income, household characteristics, marital status, and internet access.
- *Section VI—Contact Information* requested that respondents provide their personal contact information as well as contact information for two additional people who would be able to reach them in the event that they relocate.

Target Population and Estimates

Target Population

The 2004–05 TFS sample was based on interviewed public (including public charter) and private school teachers who taught students in any of grades K–12 or in comparable ungraded levels during the previous year's SASS. The sample of teachers selected included those who left the position of a K–12 teacher within the year after SASS (leavers). It also included those who continued to teach students in any of grades pre-K–12 or in comparable ungraded levels, including teachers who remained in the same school as in the previous year (stayers) and who changed schools (movers); prekindergarten was included so that sampled teachers who changed assignments from teaching students in any of grades K–12 to teaching only prekindergarten students would not be considered leavers.

In SASS, the sampling frame for public schools was an adjusted version of the 2001–02 Common Core of Data (CCD), and the sampling frame for private schools was a modified version of the 2003–04 Private School Universe Survey (PSS) sample. The sampling frame for the SASS teacher questionnaires consisted of lists of teachers provided by schools in the SASS sample. A teacher was defined as a staff member who taught a regularly scheduled class to students in any of grades K–12 or comparable ungraded levels.

Estimates

SASS was designed to produce national, regional, and state estimates for public elementary and secondary schools teachers; and national, regional, and affiliation group estimates for private school teachers.

The SASS teacher survey was designed to support comparisons between new and experienced public school teachers (3 years or less of experience vs. more than 3 years of experience) at the state level and

new and experienced private school teachers at the affiliation level. Comparisons among teachers by race and by full-time or part-time status are possible at the national level.

TFS was designed to produce national comparisons for current teachers and former teachers by school type (public or private), school level (elementary, secondary, or combined), new versus experienced, and nonminority versus minority.

Methodology

TFS is a survey that uses paper-based questionnaires that are mailed to sampled teachers for their completion. In addition to the two paper questionnaires, the 2004–05 TFS included an internet reporting option. The internet version of TFS included all of the same items from the paper questionnaires, but had automated skip patterns that presented only the appropriate questions to the respondent. An internet option was offered to a portion of the respondents with the hope that it would increase, or at least maintain, overall survey response rates while helping to reduce reporting errors by respondents in the surveys through the use of these automated skip patterns and edits.[2]

In order to draw the sample for TFS, the first step was to mail a Teacher Status Form (Form TFS-1) to all schools in which teachers completed a SASS teacher questionnaire in the 2003–04 administration of SASS. A knowledgeable person at the school, such as the principal, was asked to complete the status form by indicating the current teaching status of each teacher listed on the form. All of the teachers listed on the Teacher Status Form were sampled for SASS, but they may not have responded to the survey. If a teacher listed on the Teacher Status Form did not respond to SASS, he or she was not included on the sampling frame for TFS. The current teaching status as reported for the SASS teacher was used to determine which form, either the current or former teacher questionnaire, would be mailed to each sampled teacher. Then the Census Bureau clerical processing staff in Jeffersonville, Indiana, mailed either the current or former teacher questionnaires or letters providing an internet user name and password to the sampled teachers who had provided a home address on the SASS questionnaire. When no home address was provided by movers and leavers, these cases became part of an address research operation conducted by the Census Bureau to find a valid address. For stayers without home addresses, questionnaires or letters providing the internet user name and password were sent to their current school.

Approximately 5 weeks after the first mailout of both TFS questionnaires, respondents for which an interview was not obtained were sent a second paper questionnaire for completion. This second mailout group included those who either had not yet completed and returned their paper questionnaires or had not yet completed the internet form. An additional 5 weeks later, the cases for which a questionnaire still was not completed were sent to Census Bureau field staff who conducted follow-up using a combination of telephone interviews and personal visits.

If an interview was still not obtained during field follow-up, the case was determined to be a noninterview. TFS respondents who were deceased, had moved out of the United States, or had never been teachers (i.e., incorrectly completed the 2003–04 SASS teacher questionnaire) were determined to be out of scope for the survey. For more information about the interview status of TFS questionnaires, refer to the interview status recode (ISR) sections of chapter 6.

---

[2] The internet option was implemented as part of a test. For a complete description of the test and results, see "Appendix E. (Inter) Net Gain? Experiments to Increase Response."

Contents of the Manual

This report contains chapters on preparation for the 2004–05 TFS, frame creation and sample selection procedures, data collection, response rates, data processing, imputation procedures, weighting and variance estimation, a review of the quality of TFS data, structure of TFS data files and information on merging data files, and user notes and cautions.

Information in the chapters is supported by material in the following appendixes.

- A. Key Terms for TFS;
- B. TFS Questionnaire Availability;
- C. Summary of TFS Interview Findings and Recommendations;
- D. Summary of TFS Pretest and Usability Test Findings and Recommendations;
- E. (Inter) Net Gain? Experiments to Increase Response;
- F. Quality Assurance for TFS Keying and Mailout Operations;
- G. Changes Made to TFS Variables During Computer Edits, by Data File;
- H. Percentage of TFS Variables Changed During Three Stages of Imputation, by Data File;
- I. Weighting Adjustment Cells;
- J. Frame and Created Variables; and
- K. Crosswalk of 2004–05 TFS Items with 2000–01 TFS and 2003–04 SASS Teacher Questionnaire.

This page is intentionally left blank.

Chapter 2. Preparation for the 2004–05 TFS

The National Center for Education Statistics (NCES) and the U.S. Census Bureau continually work to improve questionnaires and procedures for the Schools and Staffing Survey (SASS) and the Teacher Follow-up Survey (TFS). Prior to the administration of the 2004–05 TFS, the survey items and methodology were tested and improved. In an effort to develop questionnaire items that would accurately capture current and former teachers' responses to the key questionnaire items, a series of cognitive interviews were conducted to identify problems that could be corrected prior to the survey's administration. The results from the first cognitive study were used to make revisions to the survey items, and a second, small-scale study was conducted to test some of these key revisions. This second study also was a usability test[3] on an internet instrument, which was added as a reporting method for the 2004–05 administration of TFS.

Cognitive Interviews and Usability Tests

Study One

In March 2004, the Census Bureau contracted with ORC Macro, a research and evaluation company in Calverton, Maryland, to carry out an initial round of cognitive interviews. The purpose of these interviews was to gather feedback from both current and former teachers on several proposed questions for the 2004–05 administration of TFS. Details on methodology and findings can be found in "Appendix C. Summary of TFS Interview Findings and Recommendations."

The test included items on

- employment outside of teaching;
- remaining in current position;
- retirement;
- decision to leave K–12 teaching/last year's position;
- last year's/current school;
- last year's/current teaching position;
- last year's/current principal or school head;
- state or district assessment exams;
- certification;
- education activities;
- returning to teaching; and
- marital status.

Methods

All cognitive interviews were conducted by trained interviewers, with each interview lasting approximately 90 minutes. Interviews were predominantly conducted on an individual basis, but some were conducted in pairs or groups, and some were conducted over the telephone. The interviews were conducted using a "think aloud" protocol, where each respondent was asked to describe what they were

---

[3] A usability test is a way to identify how users actually interact with a system (in this case, the TFS internet instrument). The goal of a usability test is to find out what is and is not working well in the instrument. In this test, respondents were instructed to describe aloud what they were thinking as they moved through the TFS internet questionnaire. Meanwhile, trained interviewers observed, asked probing questions, listened, and took notes.

thinking out loud as they completed the questions. In addition, some follow-up questions were asked after the respondent completed certain questionnaire items.

Interviews were carried out with current and former teachers in Alaska, Maine, Maryland, Massachusetts, New Hampshire, New York, and Virginia, and the District of Columbia. Respondents were offered a $60 incentive for their participation in these cognitive interviews.

Key Findings

A summary of the recommendations from the cognitive interviews follows:

- Eliminate items concerning future eligibility for retirement benefits.
- Adjust the order in which retirement items appear to make them seem less redundant.
- Eliminate items that were difficult for the respondent to answer, such as the respondent's second and third most important reasons for leaving teaching.
- Eliminate items about performance on assessments at the current school, because assessment exams may not have been administered prior to the survey in the current school year.
- Reword several questions and instructions for better clarity and uniform understanding.
- Modify answer choices to questions concerning how long the respondent plans to remain in his or her current position to include specific life events that may occur (e.g., parenthood or marriage).
- Change column title from "No difference" to "No better/worse in either position" in items that compare the current position to the former position.

Study Two

The Questionnaire for Former Teachers (Form TFS-2) and the Questionnaire for Current Teachers (Form TFS-3) were revised based upon the recommendations from the previous cognitive interview study. As a result, in July and August 2004, Census Bureau staff conducted a second series of cognitive interviews to test some of the key revisions. This second set of cognitive interviews also was a usability test of the newly developed internet instrument. Details on methodology and findings can be found in "Appendix D. Summary of TFS Pretest and Usability Test Findings and Recommendations."

The test focused on

- questionnaire content and wording—
  - o if respondents were able to interpret the questions uniformly;
  - o if respondents were able to provide the information that was requested; and
  - o if respondents were able to accurately answer the series of retirement questions; and
- internet usability—
  - o if respondents were able to log in and use the internet version of the instrument;
  - o if navigation through the instrument was clear and efficient;
  - o if respondents found it challenging to respond using the radio buttons; and
  - o if automated edits were clear, helpful, and appropriate given the respondent's answers.

Methods

All interviews were conducted by trained interviewers and were video recorded with the respondent's permission. The sample contained 24 respondents and was drawn from current and former teachers in the District of Columbia and Baltimore, Maryland, metropolitan areas. Interviews were conducted in the

Census Bureau usability lab or the respondent's home or workplace. Each respondent was offered a $50 incentive for participating in this cognitive interview study.

Key Findings

A summary of the recommendations from the pretest cognitive interviews follows:

- Questionnaire content and wording—
  o Change wording of items to provide clarity.
  o Change all items that contain the phrase "K–12 teaching" to instead say "the position of a K–12 teacher" to prevent respondent confusion.
  o Change answer formats of Individual Education Plan (IEP)/limited-English proficiency (LEP) questions from write-ins to answer choices.
  o Add specific directions to items so that the respondent could more accurately provide the requested information.
  o Change wording and merge together some of the retirement items.
- Internet usability—
  o Change the organization of the main menu to eliminate respondent confusion about starting or resuming the questionnaire.
  o Change the navigation portion of each screen, allowing respondents to advance or return to other items or sections on the questionnaire.
  o Modify dropdown boxes to make it easier for respondents to select the appropriate answer choice.

Content Changes

The TFS questionnaires were revised substantially from the 2000–01 versions prior to the testing. After both cognitive interview studies, both questionnaires were revised further based upon the results of the studies. As a result of both processes, the following additions, deletions, and revisions were made to the TFS questionnaires between the 2000–01 and 2004–05 administrations.

Changes to the Questionnaire for Former Teachers

Items Added to the Questionnaire for Former Teachers

The following topics were added to the former teacher questionnaire between the 2000–01 and 2004–05 administrations of TFS:

- maternity or paternity leave, disability leave, or sabbatical;
- supplemental income;
- retirement items;
- most important reason for leaving the position of a K–12 teacher;
- impressions of state or district assessment programs;
- enrollment in college or university courses since last school year;
- lowest teaching salary accepted to return to teaching;
- last year's marital status;
- living situation;
- internet access; and
- contact information.

Items Deleted From the Questionnaire for Former Teachers

The following topics were deleted from the former teacher questionnaire between the 2000–01 and 2004–05 administrations of TFS:

- main activity during the next school year;
- factors that influence retirement;
- years to remain in teaching if retirement were not an option;
- impressions of last year's school;
- impressions of teaching;
- degree programs and degrees earned in the past 12 months; and
- changes to teaching certification.

Items Revised on the Questionnaire for Former Teachers

The following items on the former teacher questionnaire were revised between the 2000–01 and 2004–05 administrations of TFS:

- Questions concerning current occupation were revised, and the new series does not ask for specific employer and type of business or industry.
- Question concerning how long the respondent plans to remain in current position was revised to give more options concerning retirement, retirement benefits, and specific life events.
- The section on reasons for leaving teaching was revised and includes fewer reasons.
- Questions that rate current position relative to teaching were revised to incorporate social aspects, personal life, accomplishment, and influencing the lives of others.
- Factors influencing the decision to return to teaching were revised to incorporate more factors.

Changes to the Questionnaire for Current Teachers

Items Added to the Questionnaire for Current Teachers

The following topics were added to the current teacher questionnaire between the 2000–01 and 2004–05 administrations of TFS:

- class organization;
- Individual Education Plan (IEP) and limited-English proficiency (LEP);
- general and student problems at current school;
- opinions about current school and the teaching profession;
- hours and time allocation;
- participation in school-related activities;
- grades offered in current school;
- most important reason for leaving last year's school;
- impressions of state or district assessment programs;
- rating current teaching position relative to last year's teaching position;
- enrollment in college or university courses since last school year;
- attempt to leave teaching;
- collecting pension or drawing funds from 401(k) or 403(b) plan;
- contributions to teacher retirement plan or 401(k) or 403(b) plan;
- last year's marital status;

- living situation;
- internet access; and
- contact information.

Items Deleted From the Questionnaire for Current Teachers

The following topics were deleted from the current teacher questionnaire between the 2000–01 and 2004–05 administrations of TFS:

- secondary teaching assignment fields and certifications;
- overall satisfaction with teaching in current school;
- overall satisfaction with teaching in last year's school;
- overall satisfaction with the teaching profession;
- impressions of instructional leader at current school;
- impression of administrators and staff at current school;
- impression of administrators and staff at last year's school;
- impressions of last year's school;
- planning and preparation time;
- hours and time allocation outside of teaching;
- computer use in the classroom;
- technology in the classroom;
- whether the new school is a public charter school; and
- degree programs and degrees earned in the past 12 months.

Items Revised on the Questionnaire for Current Teachers

The following items on the current teacher questionnaire were revised between the 2000–01 and 2004–05 administrations of TFS:

- Questions concerning current school conditions and experiences were revised to include fewer items.
- Reasons for leaving last year's school were revised to include fewer items.
- Question concerning how long the respondent plans to remain in teaching was revised to give more options concerning retirement, retirement benefits, and specific life events.

Final Content of the 2004–05 TFS

The following is a brief summary of the major content areas for the 2004–05 TFS. For further details about the specific sections and content of each survey, please refer to chapter 1.

- The Questionnaire for Former Teachers obtained information such as present occupation or activity, educational plans, reasons for leaving teaching, intent to return to teaching, attitudes about teaching, plans for further education, and demographic characteristics.
- The Questionnaire for Current Teachers obtained information about current teaching assignments, attitudes about teaching, attitudes about last year's school, expected duration in teaching, plans for further education, and demographic characteristics.

Copies of the 2004–05 TFS questionnaires may be obtained on the Internet at http://nces.ed.gov/surveys/sass/questionnaire.asp or by e-mail to SASSdata@ed.gov.

This page is intentionally left blank.

Chapter 3. TFS Frame Creation and Sample Selection Procedures

This chapter describes the frame creation and sampling process for the Teacher Follow-up Survey (TFS) sample. Teachers sampled for TFS were drawn from the Schools and Staffing Survey (SASS) sampled teachers who had completed interviews; those teachers were in turn drawn from the SASS sampled schools. This chapter begins with a brief description of the creation of the SASS school sampling frames. Next, the school sampling procedure is described, followed by the SASS teacher sampling process, and finally the TFS teacher sampling process.

Note that Bureau of Indian Affairs (BIA) funded schools and teachers were included in the SASS school and teacher sampling process and thus were also included in the TFS teacher frame. However, because there were so few teachers from BIA-funded schools in the TFS frame, they were dropped from the TFS sample. Therefore, the TFS sampling section of this chapter will not include any details on teachers from BIA-funded schools.

SASS Sampling Frames

Public School Sampling Frame

The public school sampling frame was based on the 2001–02 school year Common Core of Data (CCD). CCD is collected annually by the National Center for Education Statistics (NCES) from all state education agencies and is believed to be the most complete public school listing available. The frame includes traditional public schools, schools on Department of Defense military bases, BIA-funded schools, public charter schools, and nonregular schools such as special education, vocational, and alternative schools. Extensive modifications to CCD (as described below) resulted in 90,239 schools on the SASS public school sampling frame.[4]

In order to include only schools that met the definition of a school in SASS, schools were eliminated from and added to the sampling frame. In SASS, a school was defined as an institution or part of an institution that provides classroom instruction to students, has one or more teachers to provide instruction, serves students in one or more of grades 1–12 or the ungraded equivalent, and is located in one or more buildings. SASS is confined to the 50 states plus the District of Columbia and excludes territories and overseas schools. The SASS definition of a school was generally similar to CCD with some exceptions. CCD contained some schools that did not offer teacher-provided classroom instruction that included academic subjects in grades 1–12 or the equivalent ungraded levels. In some instances, schools in CCD were essentially administrative units that may have overseen entities that provided classroom instruction, or the school on CCD may have provided funding and oversight only. CCD also included schools located overseas that were operated by the Department of Defense. Consequently, schools that were ineligible for SASS and deleted from the sampling frame included schools operated overseas by the Department of Defense, adult education centers, schools in which the highest grade offered was prekindergarten or kindergarten, homeschools, tutoring services, or administrative units.

In addition, school records that appeared to have a common administration and were housed in one building were collapsed into a single school. The schools that met the criteria often offered grades K–12 in the same building or administrative unit. Because of this, these schools often perceive themselves differently than the state does (i.e., as a single entity as compared to several separate schools). For this

---

[4] For the 2001–02 school year, CCD included data on 97,623 schools.

reason, it was decided for the 2003–04 SASS to collapse the CCD records whenever it was believed that this problem was likely to occur.

SASS added schools in Pennsylvania and California after having directly contacted their administrative units as reported on CCD. These administrative units oversee certain types of educational entities (e.g., special education schools, juvenile justice facilities) within various California and Pennsylvania counties. This type of education is often provided at a number of locations within a particular county, but not necessarily at schools listed on CCD. To avoid confusion, these administrative units were contacted by phone, and requested to provide a list of the schools they oversaw. These lists were subsequently matched to CCD. If any of the schools from these lists were not already on CCD, they were added to the SASS sampling frame at that time. (For more details, see *Documentation for the 2003–04 Schools and Staffing Survey* [Tourkin et al. 2007].)

Finally, corrections and/or imputations to the school records were made to facilitate the school sampling process. These corrections included filling in and/or modifying missing grade ranges, total enrollments, enrollment by race, teacher totals, physical location components, and the school's name.

The resulting number of schools on the 2003–04 SASS public school frame was 90,239. Of these, 166 were BIA-funded schools and 2,309 were public charter schools. Additional out-of-scope schools were detected during data collection and the processing of the sampled schools' SASS school questionnaires. These schools were eliminated from further processing of the school sample and were not part of any SASS estimates of the number of schools.

Private School Sampling Frame

The sampling frame for private schools was the updated 2003–04 Private School Universe Survey (PSS) list frame updated with the 2001–02 PSS area frame. The area frame serves as a coverage improvement for the list frame.

- *List frame*. The list frame used for the 2003–04 SASS private school sample was the same list used for the 2001–02 PSS, updated in the spring of 2003 using lists from 26 private school associations and all 50 states and the District of Columbia. The resulting frame was also used for the 2003–04 PSS.
- *Area frame*. The SASS area frame consisted of a list of private schools that had not been included in the PSS universe and had not been reported by state or private school associations during the list frame updating operation. These schools were located in 116 selected primary sampling units (PSUs) throughout the United States.

Closed schools and out-of-scope schools (i.e., adult education centers, schools where the highest grade was prekindergarten or kindergarten, homeschools, or tutoring services) were deleted from the private school file before sampling. As with the public school frame, there were several corrections and/or imputations that needed to be made to school records in order for sampling to proceed. These corrections and/or imputations included modifying or filling in information for the school's grade range, affiliation, total student enrollment, and teacher counts.

The resulting number of schools on the 2003–04 SASS private school frame was 29,303 list frame schools and 179 area frame schools. As with the public schools, any additional out-of-scope schools detected during data collection or the questionnaire processing were eliminated from any SASS estimates. Thus, SASS estimates do not agree with the frame counts.

SASS School Stratification

Stratification refers to the process of subdividing the population frame into mutually exclusive subsets (called strata) from which samples of schools are selected at appropriate rates.

Public School Stratification

The first level of stratification for public and BIA-funded schools was assignment to school types A through E, as follows:

- school type A included BIA-funded schools, which were selected with certainty (automatically in sample);
- school type B included public schools with high American Indian or Alaska Native enrollment (schools with 19.5 percent or more American Indian or Alaska Native students);
- school type C included schools in Delaware, Maryland, Florida, Nevada, and West Virginia, where at least one school from each district was selected, as described in the "SASS School Sample Selection" section below;[5]
- school type D included public charter schools; and
- school type E included all other schools.

Schools falling into more than one category were assigned to types A, B, D, C, and E, in that order. For example, if a school were identified as BIA-funded as well as public charter, the school would be considered BIA-funded for stratification purposes.

The second level of stratification varied by school type, as described below:

- All type A schools were selected for the sample, so no additional stratification was needed.
- Type B schools (i.e., high American Indian or Alaska Native enrollment schools) were stratified by state, as follows: Arizona, California, Montana, New Mexico, Washington, the remaining Western states, Minnesota, North Dakota, South Dakota, the remaining Midwestern states, North Carolina, Oklahoma, and the remaining states. Note that Alaska was excluded from this group of strata because most schools in Alaska have a high Alaska Native enrollment and because the sampling rate applied to Alaska schools was higher than the sampling rate applied to other schools with high American Indian or Alaska Native student enrollment. Thus, schools in Alaska were generally included in type E.
- Type C schools (i.e., schools in Delaware, Maryland, Florida, Nevada, and West Virginia) were stratified first by state and then school district. For details, see *Documentation for the 2003–04 Schools and Staffing Survey* (Tourkin et al. 2007).
- Type D schools (i.e., public charter schools) were stratified by state, as follows: Arizona, California, Colorado, the remaining Western states, Michigan, Ohio, Wisconsin, the remaining Midwestern states, Florida, North Carolina, Texas, the remaining Southern states, Pennsylvania, and the remaining Northeastern states.
- Type E schools (i.e., all other schools) were stratified by state, as follows: all states including the District of Columbia, except those states in type C (i.e., Delaware, Maryland, Florida, Nevada, and West Virginia).

---

[5] A 2003 simulation study showed that standard errors from Delaware, Florida, Maryland, Nevada, and West Virginia were high relative to the sampling rate in SASS. To reduce the standard error, all districts in these states were defined as school sampling strata. For more details, see *Documentation for the 2003–04 Schools and Staffing Survey* (Tourkin et al. 2007).

Each of the school types, B through E, was then stratified by school level (i.e., elementary, secondary, and combined), as defined below:

- Elementary = lowest grade ≤ 6 and highest grade ≤ 8.
- Secondary = lowest grade ≥ 7 and highest grade ≤ 12.
- Combined   = lowest grade ≤ 6 and highest grade > 8, or school is ungraded.[6]

Private School Stratification

List Frame

The list frame was partitioned into an initial set of cells using affiliation stratum (17 groups), school level (3 groups), and Census region (4 groups). These cells were defined using 2001–02 PSS data. For any school records that were missing information for these three variables, the data were imputed.

The first level of stratification was school affiliation stratum (17 groups):

- Catholic—parochial;
- Catholic—diocesan;
- Catholic—private;
- Amish;
- Assembly of God;
- Baptist;
- Episcopal;
- Jewish;
- Lutheran Church—Missouri Synod;
- Wisconsin Evangelical Lutheran Synod;
- Mennonite;
- Pentecostal;
- Seventh-Day Adventist;
- other religious;
- nonsectarian—regular;
- nonsectarian—special emphasis; and
- nonsectarian—special education.

Within each affiliation stratum, schools were stratified by school level (i.e., elementary, secondary, and combined schools). The definitions are provided below:

- Elementary = lowest grade ≤ 6 and highest grade ≤ 8.
- Secondary = lowest grade ≥ 7 and highest grade ≤ 12.
- Combined   = lowest grade ≤ 6 and highest grade > 8, also includes ungraded schools.[6]

Within affiliation stratum/school level, all private schools were stratified by the four Census regions:

- Northeast (Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont);

---

[6] Ungraded schools refer to schools that serve students whose grade levels are not defined as grade 1 through 12. For example, special education centers and alternative schools often classify their students as ungraded.

- Midwest (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin);
- South (Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia); and
- West (Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming).

Area Frame

All private schools from the area frame were automatically included in the sample, so no stratification was necessary.

SASS School Sample Selection

Public School Sample Selection

To facilitate the calculation of school district weights, it was important that within a stratum all schools belonging to the same school district be listed together. This could have been achieved by sorting by the district's CCD identification number, called the Local Education Agency (LEA) ID, first. However, to increase the efficiency of the school sampling design, it was better to sort by other variables first before LEA ID. To achieve both these goals, the ZIP code variables were recoded to make them the same for every school within a stratum/school district. After the ZIP code was recoded, non-BIA-funded schools were sorted, hierarchically, by the following variables:

1. Stratum code (assigned as described in the "SASS School Stratification" section above);
2. State (one for each state and the District of Columbia);
3. Locale code—
   1 = large central city,
   2 = midsize central city,
   3 = urban fringe of large central city,
   4 = urban fringe of mid-size central city,
   5 = large town,
   6 = small town,
   7 = rural, outside Metropolitan Statistical Area (MSA), and
   8 = rural, inside MSA;
4. Recoded ZIP code (all schools in a specific stratum/district have the same value for this variable);
5. LEA ID (school district's CCD ID);
6. School's highest grade offered (in descending order);
7. Recoded percent minority (in descending order) and defined as:
   1 = less than 5.5 percent minority enrollment or unknown,
   2 = at least 5.5 percent but less than 20.5 percent minority enrollment,
   3 = at least 20.5 percent but less than 50.5 percent minority enrollment, and
   4 = at least 50.5 percent minority enrollment;
8. Total enrollment (in serpentine sort order defined as: enrollment sorted first in ascending then descending order within the other sort variables); and
9. CCD school ID.

This sort order differs slightly from the sort used in previous SASS cycles. The third and fourth variables (locale code and recoded ZIP code) allowed a geographic balance within locale at the state level to be

achieved. The fifth variable (LEA ID) guaranteed that schools within a district and school stratum would stay together. The sixth variable (school's highest grade) allowed the sample size requirements for middle schools to be met,[7] and the seventh variable (recoded percent minority) allowed a balance with respect to race/ethnicity. The eighth variable (school enrollment) also allowed a balance with respect to school size. The ninth variable, school ID, made the sort unique and therefore possible to reproduce.

Within each stratum, all non-BIA-funded schools were systematically selected for the sample from a hierarchically sorted list using a probability proportional to size algorithm. In applying this algorithm, a measure of size for each non-BIA-funded school was used to determine whether the school was selected with certainty, or whether a probability sampling procedures was applied. The measure of size, used to define the probability of selection for the schools, was the square root of the number of full-time equivalent teachers reported for each school or imputed during the sampling frame creation. Any school with a measure of size greater than the sampling interval (a measure of the spread between selected sample units in systematic sampling) was included in the sample with certainty (automatically) and excluded from the probability sampling operation. The BIA-funded schools were also selected for the sample with certainty. This produced a non-BIA sample of 10,202 (455 public schools with a high American Indian enrollment, 303 public charter schools, and 9,444 other public schools) and a BIA-funded sample of 166 schools for a total of 10,368 public and BIA-funded sample schools in 2003–04 SASS.

Private School Sample Selection

Within each stratum, sorting took place on the variables listed below. Sorting serves to improve the efficiency of the overall design.

1. State (one for each state and the District of Columbia);
2. School's highest grade offered (in descending order);
3. Locale code based on 1990 Census geography—
    1 = large central city,
    2 = mid-size central city,
    3 = urban fringe of large city,
    4 = urban fringe of mid-size city,
    5 = large town,
    6 = small town,
    7 = rural, outside MSA, and
    8 = rural, inside MSA;
4. ZIP code;
5. Enrollment as reported or imputed in the 2001–02 PSS (in descending order); and
6. PIN number (a unique number assigned to identify the school on PSS).

Within each stratum, private schools in the list frame were systematically selected using a probability proportional to size algorithm. The measure of size used was the square root of the 2001–02 PSS number of teachers in the school. Any school with a measure of size larger than the sampling interval was excluded from the probability sampling process and included in the sample with certainty. All of the area frame schools identified in the 2001–02 PSS area frame within noncertainty (selected with probability less than one) PSUs that had not already been added as part of the 2003–04 PSS list frame updating operation were also selected for the sample with certainty. This produced a list frame sample of 3,443 and an area frame sample of 179 schools, totaling 3,622 schools in the SASS private school sample.

---

[7] Since middle schools were not stratified explicitly into one grade level stratum, some of them were classified as elementary and some as secondary. For more information, see chapter 4 of *Documentation for the 2003–04 Schools and Staffing Survey* (Tourkin et al. 2007).

SASS Teacher Sample Selection

Selecting the teacher sample in both public and private schools involved the following steps:

- The selected schools were asked to provide teacher lists.
- From the lists, 53,188 teachers from public schools (including public charter and BIA-funded schools) and 9,947 teachers from private schools were selected.

The public and private teacher sample selections are described together because identical methodologies were used. The only differences were in the average number of teachers selected within a school, as shown on table 1.

SASS Teacher Frame

In previous SASS enumerations, sampled schools were asked to provide a list of their teachers primarily by mail. This cumulative list of teacher rosters formed the teacher sampling frame. For the 2003–04 SASS, sampled schools provided teacher rosters to field representatives during personal visits to the sampled schools. The field representatives keyed the roster information into a laptop computer and teachers were sampled from the school, sometimes during the same personal visit.

Along with the names of its teachers, sampled schools were asked to provide the following descriptive characteristics of each teacher:

- New/experienced—Teachers in their 1st, 2nd, or 3rd year of teaching during the 2003–04 school year were classified as new teachers, while those in their 4th year or more were classified as experienced.
- Teaching status (as defined by the school):
  - part time; or
  - full time.
- Race/ethnicity (as reported by the school):
  - White (non-Hispanic);
  - Black (non-Hispanic);
  - Hispanic—regardless of race;
  - Asian or Pacific Islander; or
  - American Indian or Alaska Native.
- Subject matter taught—Teachers were classified as teaching special education, general elementary, math, science, English/language arts, social studies, vocational/technical, or other.

The above information for each teacher in a selected SASS school comprised the teacher frame.

Within each sampled school, teachers were stratified into one of four teacher types in the following hierarchical order:

1. Asian or Pacific Islander;
2. American Indian or Alaska Native;
3. New (3 or fewer years in the teaching profession); and
4. Experienced (more than 3 years of teaching).

To illustrate the hierarchical ordering, if a teacher was both new and Asian, then that teacher would be classified as Asian.

Within-School SASS Teacher Allocation

The goals of the teacher sampling for SASS were to simultaneously achieve the following:

- Select a minimum of 1 and a maximum of 20 teachers per school.
- Select an average of three to eight teachers per school depending upon school level and school type as shown in table 1.
- Select at least 1,600 Asian or Pacific Islander teachers and 1,600 American Indian or Alaska Native teachers.
- Select a minimum of 2,300 new teachers per school type (public, private). For new teachers in public schools, oversampling was not required due to the large number of sampled schools with new teachers. Therefore, teachers were allocated to the new and experienced categories proportional to their numbers in the school. However, for private school teachers, new teachers were oversampled by a factor of 1.5. This factor was used to ensure that the sample size goal set per school type was met.
- Minimize the variance of teacher estimates within school stratum by attempting a self-weighting design (all teachers having the same probability of selection). This constraint was relaxed when necessary to accommodate the other goals of teacher sampling.

Table 1.   Average expected number of new and experienced teachers selected per school, by school level and school type: 2003–04

School type	School level		
	Elementary	Secondary	Combined
Public (including public charter and BIA-funded[1])	3.8	7.5	5.7
Private	3.8	4.7	2.8

[1] BIA refers to the Bureau of Indian Affairs.
NOTE: These averages do not include Asian or Pacific Islander or American Indian or Alaska Native teachers.
SOURCE: *Documentation for the 2003–04 Schools and Staffing Survey*, Schools and Staffing Survey (SASS), 2003–04, U.S. Department of Education, National Center for Education Statistics.

Before teachers were allocated to the new/experienced strata, schools were first allocated an overall number of teachers to be selected. This overall sample size was chosen so as to equalize the teacher weights within school stratum (i.e., state/school level for public schools and affiliation/school level/region for private schools). Teacher weights within stratum were not always equalized, however, due to the differential sampling for Asian or Pacific Islander and American Indian or Alaska Native teachers.

The final SASS teacher sample met all the goals presented above with one exception—the targeted number of American Indian or Alaska Native teachers selected was not met. Both the unweighted and weighted totals are given in the sample breakdown presented in table 2. The weighted totals are used in TFS weighting to calculate an adjustment factor. Chapter 8 gives the details of the TFS weighting process.

Table 2.   Final unweighted and weighted number of teachers selected for the SASS teacher sample, by teacher type and school type: 2003–04

			Teacher type							
	Total		Experienced		New		American Indian or Alaska Native		Asian or Pacific Islander	
School type	Unwtd.	Wtd.	Unwtd.	Wtd.	Unwtd.	Wtd.	Unwtd.	Wtd.	Unwtd.	Wtd.
Total	63,135	3,721,994	49,263	3,049,257	10,528	617,431	1,530	11,034	1,814	44,272
Private	9,947	467,373	7,008	370,368	2,496	87,169	95	1,052	348	8,784
Public charter	1,486	41,689	970	27,552	471	13,196	8	78	37	862
Traditional public and BIA-funded[1]	51,702	3,212,932	41,285	2,651,337	7,561	517,066	1,427	9,904	1,429	34,625

[1] BIA refers to the Bureau of Indian Affairs.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), 2003–04.

There were 10,368 traditional public, BIA-funded, and public charter schools selected for the sample in SASS and from these schools 53,188 teachers were selected for the sample. There were 3,622 private schools sampled in SASS and from these schools 9,947 teachers were selected for the sample. About 11 percent of public schools and 16 percent of private schools that met the SASS definition of a school did not provide teacher lists. No teachers were selected from these schools.

TFS Teacher Sampling Frame, Stratification, and Allocation

TFS Teacher Sampling Frame

The 2004–05 TFS sample consisted of 7,429 teachers out of the 51,223 public and private school teachers who participated in the 2003–04 SASS. The TFS teacher sample was selected from the pool of teachers who participated in the previous year's SASS. Similar to SASS, the 2004–05 TFS sample includes teachers from the public (including public charter) and private school sectors.

The sampling frame for TFS consisted of the 51,748 public, BIA-funded, and private school teachers who completed interviews for SASS. Any SASS teacher who did not complete an interview or was otherwise found to be out of scope for SASS was not included in the TFS frame. Teachers from BIA-funded schools were included in the sampling frame, so are included in this count. This number is slightly lower than the total number of interviewed teachers in SASS (51,847) since it excludes 99 teachers who were reported to have died or left the country at the time of the teacher status collection in the fall of 2004.

As described earlier, one purpose of TFS was to measure attrition rates a year after the 2003–04 SASS data collection. In SASS, schools were selected first and then teachers were selected within the sampled schools. The TFS teachers were then selected from the SASS eligible teacher sample.

TFS Teacher Stratification

The TFS sample is a stratified sample that was allocated in order to allow comparisons of teachers by status (stayers, movers, and leavers) within school type (traditional public, public charter, and private), experience groups, school level, and minority status. For TFS, the responding 2003–04 SASS teachers were stratified by these five variables in the order shown below:

1. School type (traditional public, public charter, private school):
   o traditional public—teachers who taught in the public school system in the 2003–04 school year;
   o public charter—teachers who taught in a public charter school in the 2003–04 school year; and
   o private—teachers who taught in a private school in the 2003–04 school year.

2. Teacher status (leaver/stayer/mover/unknown)—Each SASS sampled school was mailed a questionnaire asking for current information about the previous year's teachers. The information collected on this form was used to stratify each teacher who responded to SASS in the following categories:
   o leavers—teachers in the 2003–04 school year who left the teaching profession before the 2004–05 school year began;
   o stayers—teachers in the 2003–04 school year who remained teachers at the same school for the 2004–05 school year or teachers whose status was not reported (left blank) by the school[8];
   o movers—teachers in the 2003–04 school year who remained teachers for the 2004–05 school year but in a different school or teachers who worked in a school in the 2003–04 school year that closed or merged with another school; and
   o unknowns—teachers whose status was reported by the school as having left, without any other information given.

3. Experience (new/experienced):
   o new—teachers who had less than 4 years of teaching experience completed at the end of the 2003–04 school year; and
   o experienced—teachers who had at least 4 years of teaching experience completed at the end of the 2003–04 school year.

4. Teacher's grade level (elementary/middle/secondary):
   o elementary—teachers who taught elementary school students (any grade K–8, but at least one of grades K–4) in the 2003–04 school year regardless of the level of the school in which they taught;
   o middle—teachers who taught middle school students (grades 5–8 exclusively) in the 2003–04 school year regardless of the level of the school in which they taught; and
   o secondary—teachers who taught secondary school students (any grade 6–12, but at least one of grades 9–12) in the 2003–04 school year regardless of the level of the school in which they taught.

5. Minority status (minority/nonminority):
   o minority—teacher with any racial/ethnic background other than White, non-Hispanic; and
   o nonminority—any teacher who did not fall into the above category.

The characteristics defined above were used in all aspects of the TFS sampling. The results of this stratification are shown in table 3.

---

[8] Teachers whose status was not reported in the Teacher Status Form (Form TFS-1) were assumed to be stayers because in any given year most teachers are stayers. These teachers account for approximately 1–2 percent of all stayers.

Table 3.   TFS sampling frame counts for teachers, by school type, teacher status, teacher's grade level, teacher experience, and minority status: 2004–05

Status, grade level, experience, and minority status of teachers	Total	School type		
		Public[1]	Public charter	Private
Total	51,748	42,624	1,176	7,948
Leavers[2] (total)	4,766	3,410	247	1,109
Elementary (total)	1,614	887	93	634
New				
Minority	144	64	19	61
Nonminority	449	186	26	237
Experienced				
Minority	184	120	15	49
Nonminority	837	517	33	287
Middle (total)	981	709	41	231
New				
Minority	99	62	14	23
Nonminority	207	123	6	78
Experienced				
Minority	143	108	7	28
Nonminority	532	416	14	102
Secondary (total)	2,171	1,814	113	244
New				
Minority	144	99	13	32
Nonminority	563	431	47	85
Experienced				
Minority	240	201	14	25
Nonminority	1,224	1,083	39	102
Movers (total)	3,208	2,696	102	410
Elementary (total)	1,150	900	31	219
New				
Minority	75	57	6	12
Nonminority	340	239	12	89
Experienced				
Minority	121	105	5	11
Nonminority	614	499	8	107
Middle (total)	732	615	20	97
New				
Minority	41	36	1	4
Nonminority	201	160	5	36
Experienced				
Minority	96	83	6	7
Nonminority	394	336	8	50

See notes at end of table.

Table 3.   TFS sampling frame counts for teachers, by school type, teacher status, teacher's grade level, teacher experience, and minority status: 2004–05—Continued

Status, grade level, experience, and minority status of teachers	Total	School type		
		Public[1]	Public charter	Private
Secondary (total)	1,326	1,181	51	94
New				
Minority	84	61	14	9
Nonminority	385	324	16	45
Experienced				
Minority	132	117	9	6
Nonminority	725	679	12	34
Stayers (total)	43,774	36,518	827	6,429
Elementary (total)	14,505	10,562	394	3,549
New				
Minority	691	451	63	177
Nonminority	2,451	1,527	97	827
Experienced				
Minority	1,754	1,403	55	296
Nonminority	9,609	7,181	179	2,249
Middle (total)	8,549	7,158	137	1,254
New				
Minority	427	354	18	55
Nonminority	1,407	1,130	32	245
Experienced				
Minority	1,024	908	15	101
Nonminority	5,691	4,766	72	853
Secondary (total)	20,720	18,798	296	1,626
New				
Minority	795	681	40	74
Nonminority	3,550	3,084	78	388
Experienced				
Minority	1,981	1,826	39	116
Nonminority	14,394	13,207	139	1,048

[1] Teachers in Bureau of Indian Affairs-funded schools were included in the sampling frame as public teachers and dropped after sampling occurred.
[2] Teachers classified as having an unknown status are included in the leaver category in this table since most of them are found to be leavers.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Unpublished Sampling Frame," 2004–05.

TFS Teacher Allocation

The 2004–05 TFS sample was allocated to strata to best achieve the goals of the sampling design. To accomplish this, the following rules were applied:

- Optimize the reliability of comparisons of current versus former teachers (i.e., leavers versus nonleavers) by selecting all leavers.
- Optimize the comparison of movers versus nonmovers (stayers) by selecting all private school movers.
- Optimize the reliability of comparisons of minority versus nonminority movers.

To that end, the following procedures were applied:

- Select approximately 70 percent of the traditional public and public charter minority movers.
- Select approximately 50 percent of the traditional public and public charter nonminority movers.

Since teachers with an unknown status could be movers or leavers, sample using the mover sampling rate, as follows:

- Select all private teachers with an unknown status.
- Select approximately 50 percent of traditional public and public charter teachers with an unknown status.
- Select a fixed sample size of stayers as follows in order to optimize the comparison of stayers versus movers or leavers and to increase the number of reporting categories for publication:
  - 1,760 traditional public school stayers (approximately 5 percent of the traditional public school stayers in the sampling frame);
  - 120 public charter school stayers (approximately 15 percent of the public charter school stayers in the sampling frame); and
  - 900 private school stayers (approximately 15 percent of the private school stayers in the sampling frame).

Once the sample sizes were determined at the teacher status/school type/minority status level based on these rules, the sample was allocated to strata proportional to the cumulative measure of size (SASS teacher initial basic weight) within each stratum relative to the cumulative measure of size of the teacher status/school type/minority status level. This maximized the reliability of teacher status/school type/minority status estimates.

The final TFS sample allocation is shown below. Note that the actual selected sample was the same as the allocated sample for TFS.

Table 4.   Final allocated TFS sample sizes for teachers, by school type, teacher status, teacher's grade level, teacher experience, and minority status: 2004–05

Status, grade level, experience, and minority status of teachers	Total	School type		
		Traditional public[1]	Public charter	Private
Total	8,297	5,553	325	2,419
Leavers[2] (total)	3,582	2,330	143	1,109
Elementary (total)	1,485	786	65	634
New				
Minority	126	54	11	61
Nonminority	411	154	20	237
Experienced				
Minority	173	115	9	49
Nonminority	775	463	25	287
Middle (total)	798	539	28	231
New				
Minority	72	40	9	23
Nonminority	187	106	3	78
Experienced				
Minority	107	73	6	28
Nonminority	432	320	10	102
Secondary (total)	1,299	1,005	50	244
New				
Minority	77	40	5	32
Nonminority	319	215	19	85
Experienced				
Minority	140	108	7	25
Nonminority	763	642	19	102
Movers (total)	1,934	1,462	62	410
Elementary (total)	950	707	24	219
New				
Minority	75	57	6	12
Nonminority	285	189	7	89
Experienced				
Minority	121	105	5	11
Nonminority	469	356	6	107
Middle (total)	534	421	16	97
New				
Minority	30	25	1	4
Nonminority	135	96	3	36
Experienced				
Minority	96	83	6	7
Nonminority	273	217	6	50

See notes at end of table.

Table 4.   Final allocated TFS sample sizes for teachers, by school type, teacher status, teacher's grade level, teacher experience, and minority status: 2004–05—Continued

Status, grade level, experience, and minority status of teachers	Total	School type		
		Traditional public[1]	Public charter	Private
Secondary (total)	450	334	22	94
New				
Minority	41	25	7	9
Nonminority	138	88	5	45
Experienced				
Minority	61	49	6	6
Nonminority	210	172	4	34
Stayers (total)	2,781	1,761	120	900
Elementary (total)	1,315	772	71	472
New				
Minority	112	68	12	32
Nonminority	437	246	18	173
Experienced				
Minority	102	68	9	25
Nonminority	664	390	32	242
Middle (total)	639	451	25	163
New				
Minority	58	45	4	9
Nonminority	206	148	6	52
Experienced				
Minority	46	36	2	8
Nonminority	329	222	13	94
Secondary (total)	827	538	24	265
New				
Minority	58	42	3	13
Nonminority	297	190	7	100
Experienced				
Minority	53	39	3	11
Nonminority	419	267	11	141

[1] Teachers in Bureau of Indian Affairs-funded schools were included in the allocation.
[2] Teachers classified as having an unknown status are included in the leaver category in this table since most of them are found to be leavers.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "TFS Sample File," 2004–05.

TFS Teacher Sample Selection

Sorting

Within each TFS stratum, teachers who had completed interviews (i.e., had a final interview status recode [ISR] = 1) in the 2003–04 SASS were sorted by measure of size, subject taught, Census region, affiliation membership (private teachers only), school locale, school enrollment, and SASS teacher control number to achieve a random, balanced sample. The variables used in the sort are described below:

1. Measure of size—this is the 2003–04 SASS teacher initial basic weight (inverse of the probability of selection prior to any corrections identified during data collection).

2. Recoded teacher subject (based on SASS teacher responses)—this is the main subject that a teacher taught during the 2003–04 school year:
   o special education;
   o general elementary;
   o math;
   o science;
   o English/language arts;
   o social studies;
   o vocational/technical; or
   o other.

3. Census region—the region in which the school is located:
   o Northeast (Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont);
   o Midwest (Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin);
   o South (Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia); or
   o West (Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming).

4. Recoded affiliation stratum (private school teachers only, based on the SASS private school stratification)—the affiliation with which the school is associated, including:
   o Catholic—parochial;
   o Catholic—diocesan;
   o Catholic—private;
   o Amish;
   o Assembly of God;
   o Baptist;
   o Episcopal;
   o Jewish;
   o Lutheran Church—Missouri Synod;
   o Wisconsin Evangelical Lutheran Synod;
   o Mennonite;
   o Pentecostal;
   o Seventh Day Adventist;
   o other religious;

   o nonsectarian—regular;
   o nonsectarian—special emphasis; or
   o nonsectarian—special education.

5. Locale based on 1990 Census geography—that is, the area in which the school is located. The categories are as follow:
   o large central city;
   o mid-size central city;
   o urban fringe of large central city;
   o urban fringe of mid-size central city;
   o large town;
   o small town;
   o rural, outside MSA; and
   o rural, inside MSA.

6. School enrollment—that is, the number of students enrolled in the school during the 2003–04 school year.

7. Teacher control number—that is, the number assigned to each 2003–04 SASS sampled teacher.

Sample Selection

After the teachers were sorted using the above variables they were selected within each stratum using a systematic probability proportional to size sampling procedure. This procedure was similar to that used in the SASS school selection. Any teacher with a measure of size (SASS teacher initial basic weight) greater than the sampling interval was included in the sample with certainty (automatically included). Since TFS selection probabilities were not conditioned on anything, the selected sample sizes equaled the allocated sample size. At this point, teachers from BIA-funded schools were dropped from the TFS sample.[9]

---

[9] There were 40 teachers from BIA-funded schools who were deleted from the TFS sample.

This page is intentionally left blank.

Chapter 4. Data Collection

The U.S. Census Bureau conducted the 2004–05 Teacher Follow-up Survey (TFS) during the 2004–05 school year. TFS data collection began with teacher sampling procedures; schools were mailed a Teacher Status Form to complete by indicating the occupational or teacher status of each teacher listed on the form. If a school did not respond to the Teacher Status Form, those teachers were still eligible for sampling in TFS. Next, a sample of teachers was mailed a Questionnaire for Former Teachers or a Questionnaire for Current Teachers to complete. A portion of the sampled teachers was given the option to complete the questionnaire on the Internet. Finally, follow-up efforts were conducted to collect TFS questionnaire data from those sampled teachers who had not yet completed a questionnaire.

Overview of Data Collection Procedures

TFS data collection began as part of the sample selection operation in the fall of 2004. The Teacher Status Form was mailed to each school that had at least one teacher who participated in the 2003–04 Schools and Staffing Survey (SASS). A knowledgeable person at the school (e.g., a school administrator, a member of the office staff) was asked to complete the Teacher Status Form by indicating the current teaching or other occupational status of each teacher listed on the form. The sample for TFS was selected based on the status of teachers obtained in the Teacher Status Form operation.[10] For additional information about the sample selection procedures for TFS, see chapter 3.

Because paper questionnaires were the primary data collection method for the 2004–05 TFS, valid addresses were needed to distribute the questionnaires to respondents. The addresses that typically were used were home addresses that the respondent provided on the 2003–04 SASS teacher questionnaire. However, because some respondents did not provide a home address on SASS, Census Bureau clerical staff conducted an initial address research operation in order to obtain a valid home mailing address so that the questionnaires could be distributed. A similar operation was used for cases that were returned by the post office as "undeliverable as addressed."

In January 2005, as part of an experiment on response rates for mailed vs. internet questionnaires, approximately one-third of the selected teachers were mailed paper questionnaires. The remaining respondents were offered the option to complete the questionnaire on the Internet and were mailed a letter that included the user name and password that they could use to access the survey on the Internet. At the end of January, these internet respondents were also mailed paper questionnaires. The internet option was offered in the context of an experiment that included providing monetary incentives to roughly half of the respondents in order to increase response rates, specifically internet responses. With increased response rates, the number of cases sent to field for follow-up would be reduced.

After the initial and follow-up mailing of questionnaires, approximately 58 percent of respondents had not completed the TFS questionnaire on paper or on the Internet. Census Bureau field staff was responsible for completing data collection for these cases. As fieldwork progressed, 6 percent of the questionnaires in the initial field workload were mailed in late by respondents who had completed them on their own before field staff had a chance to contact them. These cases were removed from the field staff workload. Field staff contacted all nonrespondents, and this resulted in a 70 percent response rate for cases assigned to field. Supervisors at 12 Regional Offices coordinated the field staff's work. Regional Office staff was responsible for making assignments, supervising fieldwork, and sending completed questionnaires to the processing staff.

---

[10] The focus primarily was to distinguish between teachers still teaching at the same school as in the 2003–04 school year, teachers who moved to new schools, and teachers no longer teaching in grades pre-K–12.

The Census Bureau processing staff in Jeffersonville, Indiana, was responsible for checking-in completed questionnaires, capturing data, and implementing quality control procedures.

An overview of the purpose and content of each TFS questionnaire is discussed in chapter 1. The preparation for the 2004–05 TFS is described in chapter 2.

Internet and Incentive Experiment

The goal of the 2004–05 TFS internet and incentive experiment was to use monetary incentives to increase overall response rates and responses via the Internet, when both mail and internet choices were offered. The current and former teachers who were selected to be in the TFS sample were randomly divided into six different treatment groups. The groups varied on three dimensions, which included offering an internet option, offering a prepaid $10 incentive, and notifying respondents in the internet groups that they also would receive a paper questionnaire.

There was an initial group of movers and leavers who did not provide a home address on the 2003–04 SASS and therefore could not be mailed a TFS questionnaire. Then, as TFS operations progressed, additional cases were returned to clerical processing staff with inaccurate or unreachable addresses. These cases were all sent to address research operations and were removed from the internet and incentive experiment. For more information about this operation, see the "Address Research Operations" section of this chapter. The resulting sample sizes for the TFS internet and incentive experiment can be found in table 5.

Groups 1 and 2 were sent only paper questionnaires throughout the duration of the experiment. Group 1 was given a $10 incentive card with the first mailing of TFS materials, and group 2 was not given any sort of incentive. Groups 1 and 2 were included in the experiment so that Census Bureau analysts could evaluate the impact of the incentive on mail-only questionnaires.

Groups 3 through 6 initially were given the internet option and shortly afterwards were given the paper option as well. However, groups 3 and 4 were not made aware that they would receive a paper version of the TFS a week later, while groups 5 and 6 were told of the forthcoming paper questionnaire in the initial letters they received requesting their participation. Finally, the four internet groups were further broken down into incentive panels. Groups 3 and 5 were given a $10 incentive card with the first mailing of TFS materials. The remaining groups were not offered any kind of monetary incentive. Table 5 below gives the characteristics of the six treatment groups.

Table 5.   Treatment groups, by distribution mode, incentive offered, and sample size: 2004–05

Group and distribution mode	Incentive offered	Sample size
1: No Internet	Yes	1,074
2: No Internet	No	1,147
3: Internet, **without** mention of mailout 1 week later	Yes	1,096
4: Internet, **without** mention of mailout 1 week later	No	1,100
5: Internet, **with** mention of mailout 1 week later	Yes	1,067
6: Internet, **with** mention of mailout 1 week later	No	1,131

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

At the beginning of the experiment, all teachers were mailed either a letter offering the internet option (groups 3–6) or a letter and questionnaire at the same time (groups 1–2). The $10 incentive card was included in this mailing for groups 1, 3, and 5. The incentive card was a prepaid $10 American Express

gift card that could be used wherever it was accepted. The cards expired approximately 6 months from when they were received by the sampled teachers. Respondents who received the incentive were given an 800-number to call if they had any problems using the card. Respondents offered the internet reporting option were given an 800-number as well as an e-mail address to assist with any questions or problems they encountered.

Key findings from the 2004–05 TFS internet and incentive experiment can be summarized as follows:

- The best response to TFS was achieved by offering only the mailed paper questionnaires.
- While adding an internet option negatively impacted the total number of responses, the negative effect was offset by not mentioning the forthcoming mail option.
- Offering a small, prepaid incentive increased responses by 5.5 percent, which is statistically significant ($p < .05$).

For a detailed description of the background, methodology, and results of the TFS internet and incentive experiment, refer to "Appendix E. (Inter) Net Gain? Experiments to Increase Response."

Timing of Data Collection

Data collection for the 2004–05 TFS began in October 2004 and continued through June 2005. Table 6 summarizes the specific data collection activities and the time frame in which each occurred.

Table 6. Data collection time schedule: 2004–05

Activity	Date of activity
Teacher Status Forms and letters mailed to sampled schools	Oct. 2004
Reminder postcards for the Teacher Status Form mailed to sampled schools	Oct. 2004
Nonresponse follow-up of schools that did not return the Teacher Status Form	Nov. 2004
Address research operation (before mailout)	Jan. 2005
Initial mailing of paper current and former teacher questionnaires to non-internet groups; initial mailing of letter containing user name and password to access survey on the Internet to internet groups	Jan. 2005
Mailing of current and former teacher questionnaire reminder postcards to non-internet groups	Jan. 2005
Initial mailing of paper current and former teacher questionnaires to internet groups	Jan. 2005
Initial mailing of current and former teacher questionnaires to address research cases	Feb. 2005
E-mail reminders sent to respondents with valid e-mail addresses from the Teacher Status Form	Feb. 2005
Telephone and personal visit follow-up for cases with no success in the address research operation	Feb.–Mar. 2005
"Undeliverable as addressed" address research operation	Feb.–Mar. 2005
Second mailing of current and former teacher questionnaires to both internet and non-internet groups	Mar. 2005
Mailing of current and former teacher questionnaires to "undeliverable as addressed" address research cases	Mar. 2005
Telephone and personal visit follow-up of current and former teacher questionnaire nonrespondents	Apr.–Jun. 2005

SOURCE: *Documentation for the 2003–04 Teacher Follow-up Survey*, Teacher Follow-up Survey (TFS), 2004–05, U.S. Department of Education, National Center for Education Statistics.

Data Collection Procedures

Collecting Teacher Status Information

In October 2004, the Census Bureau's clerical processing staff mailed the Teacher Status Form to sampled schools that provided lists of teachers during the 2003–04 SASS. The schools were asked to complete the form by indicating whether each teacher listed was still teaching in that school (stayer), was teaching in another school (mover), or had left the position of a K–12 teacher (leaver). The status of the teacher was needed to select the sample for TFS. The Teacher Status Form also requested an e-mail address for each current or former teacher listed; reminders were sent to these addresses later in the data collection phase.

One week after the Teacher Status Form mailing, reminder postcards were sent to the sampled schools. After the Teacher Status Forms and reminder postcards were distributed to schools, approximately 49 percent of schools had not completed a Teacher Status Form. Census Bureau clerical staff was responsible for the nonresponse follow-up of these cases. Nonresponse efforts consisted of staff following a script to make telephone calls to the schools in order to obtain teacher status information. Staff documented each call attempt by entering an outcome code in a call record; this outcome code indicated what had happened during each follow-up attempt (e.g., a complete interview was collected, a partial interview was collected, the school refused to participate, etc.). A handful of nonresponse cases were Amish schools that did not have telephone numbers and, therefore, could not be contacted by clerical staff during nonresponse follow-up. Census Bureau staff conducted the follow-up with personal visits for these cases. Of the 28 cases assigned to staff, 25 interviews were completed. The final response rate for the Teacher Status Form operation was 99 percent. For further information about the Teacher Status Form operation and TFS sample selection procedures refer to chapter 3.

Initial Mailings of Questionnaires to Current and Former Teachers

In January 2005, the first mailing of questionnaires took place. All teachers were mailed a version of the TFS introductory letter that explained the purpose of the survey and included a statement of authority and assurance of confidentiality. Each of the six treatment groups had its own version of the letter. The respondents in the internet groups (groups 3–6) were provided a user name and password to use to access the survey on the Internet in this initial letter. The respondents in the non-internet groups (groups 1–2) received paper versions of either the Questionnaire for Former Teachers or the Questionnaire for Current Teachers with the letter. Finally, the respondents in the incentive groups (groups 1, 3, and 5) received their $10 incentive card with the letter.

The non-internet groups were mailed a reminder postcard approximately 10 days after the first mailing in January 2005. At the same time, the internet groups were mailed the paper current or former teacher questionnaires and were reminded about the internet option. In February 2005, e-mail reminders were sent to respondents whose e-mail addresses were provided on the Teacher Status Form. The e-mail reminders sent to respondents in the internet groups included a link to the TFS questionnaire website.

The Questionnaire for Former Teachers was sent to sampled persons who were reported by school administrators as having left the K–12 teaching profession at the home address that was provided in the 2003–04 SASS teacher questionnaire. The Questionnaire for Current Teachers was sent to sampled persons who were reported as still teaching at the elementary or secondary level, either in the same school as the previous year (stayers) or in a different school (movers). Questionnaires were sent to the stayers' home address when one was provided on SASS. Otherwise, they were sent to the stayers' school address.

Current teacher questionnaires were sent to the movers' home address that was provided on SASS, if one was provided.

Address Research Operations

Movers and leavers who did not provide a home address on the 2003–04 SASS could not be mailed a questionnaire. Therefore, in January 2005, clerical staff began an address research operation. Staff began the operation by calling the sampled teacher's contact person(s). On the 2003–04 SASS teacher questionnaires, teachers were asked to provide the address and telephone number of two people who would know how to get in touch with them during the coming years. If the sampled teacher did not provide a contact person(s) on SASS, or the contact person(s) was not helpful, staff called the former school and used online databases to try to obtain new contact information for the sampled person. This included a new home address, work address, and/or telephone numbers at both. If a new address of some sort was found, the questionnaire was sent to the new address in February 2005 to be completed by the sampled teacher. If no new address was found, the questionnaire was labeled with his or her former school address and sent to one of the 12 Regional Offices[11] to be assigned to field staff. Investigative follow-up for these cases took place throughout February and March 2005. Staff contacted the schools where the teachers worked the year before to try to obtain location information. All cases included in the address research operation were removed from the incentive experiment because data collection began later for this group, making the collection process incomparable to cases included in the incentive experiment.

Throughout February and March 2005 and after the first mailing, cases that were returned to the Census Bureau's clerical processing center as "undeliverable as addressed" by the post office were sent to clerical staff for address research. If a new address was found, the questionnaire was sent to the new address in March 2005 to be completed by the sampled teacher. If no new address was found, the questionnaire was addressed to the teacher at his or her former school and sent to one of the 12 Regional Offices to be assigned to field staff with the nonresponse cases. These cases were removed from the TFS incentive experiment analysis.

Follow-up Mailings

In March 2005, approximately 6 weeks after the first mailing, each sampled person who had not completed a TFS questionnaire, either on the paper form initially mailed to them or on the Internet, was mailed a second questionnaire. The internet groups were once again reminded about the internet option.

"Switchers"

Throughout the data collection period, many respondents returned the questionnaire they received indicating that it did not apply to them. This happened when the sample teacher's 2003–04 school incorrectly reported the sample teacher's current teaching status on the Teacher Status Form. These cases were referred to as "switcher" cases. For example, if sampled teacher A's 2003–04 school reported her as leaving the teaching profession, she was categorized as a "leaver" and mailed a Questionnaire for Former Teachers. However, it may have been the case that this teacher left her previous (2003–04) school and was working as a teacher elsewhere. This would make her a "mover," meaning that she was currently teaching and should complete the Questionnaire for Current Teachers rather than the Questionnaire for Former Teachers that she was mailed originally. Teacher A is called a "switcher."

---

[11] The Census Bureau has Regional Offices in Atlanta, Boston, Charlotte, Chicago, Dallas, Denver, Detroit, Kansas City, Los Angeles, New York, Philadelphia, and Seattle.

In the case of switchers, the sampled teacher, Teacher A, was instructed to return the incorrect form that she received (in this case, a Questionnaire for Former Teachers) to clerical processing staff who would then send her the correct form (a Questionnaire for Current Teachers) for completion. Among TFS cases, 4.6 percent were switchers.

The internet version contained both versions of TFS. Respondents' answers to the initial screening questions determined if they followed the former or current teacher questionnaire path. In addition, the internet instrument automated skip patterns and presented the appropriate questions based on the respondent's answers.

Nonresponse Follow-up

Follow-up efforts began in April 2005. Cases were included in nonresponse follow-up when the sampled teachers had not returned a completed paper questionnaire or completed the questionnaire on the Internet. The majority of the nonresponse cases had valid telephone numbers and home addresses, so field staff contacted teachers by telephone and personal visit to complete the TFS interview. For the remaining cases, nonresponse efforts consisted of telephone calls to all sources of information available—contact persons from SASS, information found during the address research operation, and the prior year's school—in order to obtain information about the person's current home address or employer. If at least four attempts were made by telephone and no contact was established or there was no valid telephone number or no way to obtain a valid one, personal visits were permitted. Field staff members were instructed to visit the former schools and attempt to obtain information to locate the sampled persons.

Field staff members were supplied with a preprinted labeled questionnaire for each nonresponse case. The questionnaires were labeled with any and all available information, including home address, former or current school address, contact information, and any information that was obtained during the address research operations, where applicable. In addition, they were instructed to have blank former and current teacher questionnaires on hand at all times in case they encountered "switcher cases" that required the other type of questionnaire.

After the field staff member completed a follow-up action (e.g., conducted an interview, picked up a completed questionnaire, left a message, verified that the questionnaire had been mailed), he or she indicated what had occurred by entering an outcome code and notes into the TFS Excel spreadsheet, which kept track of the status of all TFS nonresponse follow-up cases. The Regional Offices mailed all completed questionnaires to the clerical processing staff weekly. Progress reports for the nonresponse cases were run twice weekly. Final response rates are presented in detail in chapter 5.

Information about questionnaire check-in, data capture methods used to convert data from paper to electronic format, and criteria for determining final response rates can be found in chapter 6.

Changes in Data Collection Methods from the 2000–01 TFS

A number of new data collection methods were implemented in the 2004–05 administration of TFS. These methods included a series of address research operations, an internet reporting option, the offering of monetary incentives to respondents, and the use of e-mail reminders to respondents. Each of these methods was employed with the ultimate goal of increasing overall response to TFS.

Address Research Operation

In the 2000–01 TFS, all cases without addresses were sent directly to field staff, who called the contacts provided by the teachers in SASS (if they were provided) or the teachers' former schools. The 2004–05 TFS conducted a centralized research operation for these cases, which included researching addresses in an extensive U.S. address database, in addition to calling the contacts that the sampled teacher provided on the 2003–04 SASS teacher questionnaire and the sampled teacher's 2003–04 school.

Internet and Incentives

The 2000–01 TFS questionnaires were distributed and data were collected by mail only. However, for the 2004–05 TFS, an internet reporting option was implemented for two-thirds of the respondents. Census Bureau analysts and the National Center for Education Statistics (NCES) wanted to encourage responses on the Internet because internet questionnaires can reduce nonsampling error due to the automated skip patterns, edits, and range checks that are embedded in internet questionnaires. Additionally, internet questionnaires can provide data processing advantages such as faster turnaround times and lower processing costs.

While one of the most important goals of TFS was to maximize the overall response rates, simply offering an internet reporting option typically does not increase total response rates for mailed questionnaires.[12] Census Bureau analysts and NCES wanted to implement an additional data collection method that would increase internet response (in addition to mail response). Consequently, a $10 monetary incentive was offered to approximately half of the TFS respondents. By offering an internet option as well as a monetary incentive, analysts hoped that the overall response rate would be higher than it would be if these methods were not employed. This was tested in the 2004–05 internet and incentive experiment, which is described further in "Appendix E. (Inter) Net Gain? Experiments to Increase Response."

E-mail Reminders

The Teacher Status Forms that were mailed to schools requested an e-mail address, in addition to teaching or other occupational status, for each of the teachers listed. Later during data collection, respondents for whom an e-mail address had been collected were sent an e-mail reminder. These e-mail reminders were sent prior to the second questionnaire mailings and served as a way of reminding sampled teachers that they should complete the appropriate TFS questionnaire. In addition, the reminders e-mailed to respondents who received the internet reporting option contained a link directly to the internet survey, making it easy for respondents to get to the survey. Census Bureau analysts and NCES hoped that reminding respondents of TFS using another mode of communication (i.e., e-mail) would help to increase the self-administered response on the survey.

---

[12] Several studies have obtained this result. See "Appendix E. (Inter) Net Gain? Experiments to Increase Response."

This page is intentionally left blank.

Chapter 5. Response Rates

This chapter presents the overall survey, or unit, and item response rates for the 2004–05 Teacher Follow-up Survey (TFS). The unit response rates for each TFS survey are presented in detail, and the item response rates for both the Questionnaire for Former Teachers and the Questionnaire for Current Teachers are summarized. Nonresponse bias analyses were conducted on both the unit (i.e., type of TFS questionnaire, meaning the current teacher or former teacher questionnaire) and the individual items for TFS. These analyses are described and the major findings are presented following the response rate sections.

Survey Response Rates

Sampled cases fall into one of three categories: a completed interview, a noninterview, or out of scope. A completed interview means that a sampled teacher who met the criteria for inclusion[13] in TFS substantially completed[14] the appropriate TFS questionnaire (i.e., the former teacher questionnaire or the current teacher questionnaire). Noninterviews refer to sampled teachers who met the criteria for inclusion in TFS, but did not complete the questionnaire. Out-of-scope cases were deemed ineligible to participate in TFS and were not included in the TFS sample. SASS respondents were designated as out of scope if they moved out of the United States following the 2003–04 school year, were deceased, or had never been teachers (i.e., incorrectly reported their teaching status in the 2003–04 Schools and Staffing Survey [SASS]).

A unit response rate is the rate at which the sampled units responded by sufficiently completing the questionnaire. Unit response rates can be calculated as unweighted or weighted. The unweighted response rates are the number of interviewed cases divided by the number of eligible sampled units (i.e., including interviews and noninterviews, but not out-of-scope cases). The weighted response rates are the base-weighted (initial basic weight multiplied by the sampling adjustment factor) number of interviewed cases divided by the base-weighted number of eligible cases. The initial base weight for each sampled unit is the inverse of the probability of selection. For further discussion of the weighting procedures followed for the 2004–05 TFS, see chapter 8.

Tables 7 and 8 summarize the weighted and unweighted response rates for cases in the 2004–05 TFS by data file and by the sector of the teacher's base-year school (i.e., public or private). The response rate tables are useful as indicators for possible nonresponse bias. The unweighted response rates provide an indication of the general success of the data collection efforts, while the weighted response rates provide measures of the quality of the data and the potential for nonresponse bias.

The weighted unit response rate for all teachers in the 2004–05 TFS was 91.82 percent.[15] Table 7 summarizes the weighted and unweighted response rates for cases in the 2004–05 TFS by data file (i.e., current and former teachers) and by the sector of the teacher's base-year school (i.e., public or private). The response rate for current teachers (shown in tables 7 and 8) includes teachers who stayed in the same school for the 2004–05 school year (stayers) and those who moved to a new school (movers). Both stayers and movers completed the current feacher questionnaire. The weighted response rate for movers

---

[13] Sampled teachers met the criteria for inclusion in the TFS if they completed a 2003–04 Schools and Staffing Survey teacher questionnaire and were living in the United States.

[14] To be classified as a completed interview, the respondent had to answer three required questions on the Former Teacher Questionnaire and four required questions on the Current Teacher Questionnaire and at least 10 percent of the remaining questionnaire items. (See the "Final Interview Status Edit" section in chapter 6 for more details.)

[15] For the formula to calculate the unit response rate, see Standard 1-3 in the *NCES Statistical Standards* (Seastrom 2003).

(90.91 percent) was lower than the weighted response rate for stayers (92.11 percent), reducing the response rate for current teachers to 92.03 percent. The response rate for former teachers who completed the former teacher questionnaire was slightly lower at 89.64 percent.

Table 7.   Unweighted and base-weighted response rates (in percent) of sampled teachers, by sector and teaching status: 2004–05

Sector and teaching status	Unweighted response rate	Base-weighted response rate
Total	91.02	91.82
Current teachers	92.17	92.03
Former teachers	89.95	89.64
Public[1]	92.03	91.90
Current teachers	92.49	92.06
Former teachers	91.16	90.25
Private	88.33	91.14
Current teachers	90.96	91.86
Former teachers	84.56	86.58

[1] The public sector includes teachers from traditional public and public charter schools.
NOTE: Weighted response rates use the inverse of the probability of selection and the sampling adjustment factor.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

The overall response rate represents the response rate to the survey taking into consideration each stage of data collection. For a teacher to be eligible for TFS, it was necessary to have received the Teacher Listing Form from the school during the 2003–04 SASS data collection, which provided a sampling frame for teachers at that school, and for the teacher to have responded to the SASS teacher questionnaire. This overall response rate is the product of the survey response rates shown in table 8: (SASS Teacher Listing Form response rate) x (SASS teacher questionnaire response rate) x (TFS questionnaire response rate).

Table 8.   Base-weighted response rates (in percent) for SASS teacher data files and TFS data files, by sector: 2003–04 and 2004–05

Sector	Base-weighted 2003–04 SASS Teacher Listing Form response rate	Base-weighted 2003–04 SASS teacher data file response rate	Base-weighted 2004–05 TFS response rate		Overall 2004–05 TFS response rates	
			Current teachers	Former teachers	Current teachers	Former teachers
Total	88.27	84.55	92.03	89.64	68.68	66.90
Public[1]	89.20	84.90	92.06	90.25	69.72	68.35
Private	85.40	82.40	91.86	86.58	64.64	60.92

[1] The public sector includes teachers from traditional public and public charter schools.
NOTE: Base-weighted response rates use the inverse of the probability of selection and the sampling adjustment factor.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public and Private School Teacher Documentation Data Files," 2003–04; Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

Item Response Rates

Item response rates indicate the percentage of respondents that answered a given survey question, or item. The weighted TFS item response rates are produced by dividing the number of sampled teachers who responded to an item by the number of sampled teachers who were eligible to answer that item, adjusting by the final weight. For all TFS items, a counted response is any item that is not missing and that has a value of "0" for the associated imputation flag. For detailed information on imputations performed on TFS datasets, see chapter 7.

The weighted item response rates for the former teacher questionnaire ranged from 41 percent to 100 percent. The weighted item response rates for the current teacher questionnaire ranged from 63 percent to 100 percent. Table 9 provides a brief summary of the item response rates for both questionnaires. The item response rates in these tables are weighted and do not reflect additional response loss due to respondents' refusal to participate in the survey.

Table 9. **Summary of weighted item response rates for the Teacher Follow-up Survey, by questionnaire: 2004–05**

Questionnaire	Range of item response rates (percent)	Percentage of items with a response rate of 85 percent or more	Percentage of items with a response rate of 70.0–84.9 percent	Percentage of items with a response rate of less than 70 percent
Questionnaire for Former Teachers	40.81–100.00	94.00	2.00	4.00
Questionnaire for Current Teachers	63.00–100.00	95.85	3.55	0.60

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

On the former teacher questionnaire, there were six items that had a weighted response rate of less that 85 percent. On the current teacher questionnaire, there were seven items that had a weighted response rate of less than 85 percent. These items are described in exhibit 1.

Exhibit 1. **Items with weighted response rates of less than 85 percent, by questionnaire: 2004–05**

Questionnaire	Item
Questionnaire for Former Teachers	7 (How much)
	26
	26 Yes (1 or 2)
	31
	35
	35 (Other)
Questionnaire for Current Teachers	3
	5 (Main assignment)
	4a
	15
	20 (ZIP code)
	20 (School district)
	44

NOTE: Numbers in this exhibit refer to questionnaire item numbers, while letters or parenthetical descriptions refer to subitems.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

Nonresponse Bias Analysis

A comprehensive nonresponse bias analysis was conducted for the 2004–05 TFS. The analysis evaluated the extent of potential bias introduced by teacher nonresponse at both unit and item levels.

Unit-level Nonresponse

Overview of Methodology

The first step in conducting the bias analysis was to examine the overall response rate by the reporting characteristics (i.e., urbanicity, school level, teacher status, teaching experience, and race/ethnicity). If the response rate fell below 50 percent, that population would not be reported separately in a published table; instead, the data would be replaced with a double dagger, although the estimates would be included in the total. The footnote would read, "Reporting standards not met. The base-weighted unit response rate was below 50 percent." For any characteristic where the response rate was less than 85 percent, a more detailed analysis was done on the characteristics used in stratification. The results were highlighted if that particular cell had a significantly higher or lower response rate than the file as a whole and bolded if the difference was noteworthy. A noteworthy difference had to meet the following conditions:

- The difference relative to the overall response rate, or frame proportion, was greater than 10 percent.
- The absolute difference was greater than one percentage point.
- The coefficient of variation was less than 15 percent.
- The cell had at least 30 interviews.

In addition, the base-weighted distribution of the respondents was compared to the distribution on the frame, which was adjusted for sampled units identified as out of scope. As discussed above, significant differences were highlighted and noteworthy cells were bolded.

Comparing the overall response rate of each file to the tabulation cells helped to identify areas of potential concern. Comparing the base-weighted distribution of the respondents to the adjusted frame helped to identify areas of potential bias for data items that were not particularly well correlated with the weighting cells.

Summary of Conclusions. Evidence of substantial bias was not found on the data files.

Summary for TFS Teachers

The overall response rate for the teachers was 91.8 percent. The more detailed analysis was performed by the primary reporting characteristics (i.e., urbanicity, teacher's grade level, school type, teaching experience, teacher status, school size, percent minority enrollment, teacher's age, teacher's race/ethnicity, and region). The overall response rate for Hispanic teachers was below 85 percent. No other group had a noteworthy response rate below 85 percent. The response rate breakdown is listed in table 10.

Table 10.   Number of teachers in sample and base-weighted unit response rates, by selected reporting characteristics: 2004–05

Reporting characteristic	Number of eligible teachers	Response rate (percent)	Standard error (percent)	*t* statistic
Total	8,168	91.8	0.55	†
All teachers in sample				
Urbanicity				
Central city	2,495	91.4	1.15	-0.3054
Suburban	3,794	90.9	0.88	-0.9297
Rural	1,879	94.5	1.00	2.3600
Teacher's grade level				
Elementary	3,681	92.1	0.82	0.3069
Middle	1,941	92.2	1.21	0.2993
Secondary	2,546	91.0	1.15	-0.6292
School type				
Traditional public	5,463	91.9	0.66	0.0872
Public charter	321	92.7	2.37	0.3472
Private	2,384	91.1	2.04	-0.3191
Teaching experience				
New	3,024	91.0	0.90	-0.7744
Experienced	5,144	92.1	0.70	0.2652
Teacher status				
Leaver	1,588	93.0	1.37	0.7839
Mover	1,903	91.5	1.17	-0.2276
Stayer	2,771	92.0	0.66	0.2021
Unknown	1,906	89.0	1.62	-1.6342
School size				
Less than 200	1,699	91.8	2.73	-0.0192
200–349	1,299	93.1	1.41	0.8666
350–499	1,347	91.9	1.68	0.0488
500–749	1,490	91.9	1.09	0.0449
750 or more	2,333	91.3	0.96	-0.5090
Percent minority enrollment				
Less than 10 percent	2,678	91.8	1.07	-0.0127
10–34 percent	2,315	93.1	1.25	0.9638
35 percent or more	3,175	90.9	0.90	-0.8666
Teacher's age				
Less than 30	1,938	89.7	1.68	-1.1836
30–39	1,944	90.7	1.31	-0.7913
40–49	1,619	93.4	0.96	1.4428
50–64	2,511	92.5	0.98	0.5724
65 or more	156	92.1	4.14	0.0673

See notes at end of table.

Table 10.   Number of teachers in sample and base-weighted unit response rates, by selected reporting characteristics: 2004–05—Continued

Reporting characteristic	Number of eligible teachers	Response rate (percent)	Standard error (percent)	*t* statistic
All teachers in sample—Continued				
Teacher's race/ethnicity				
Hispanic, any race	395	82.8	3.27	-2.7216
Asian or Pacific Islander, non-Hispanic	259	90.7	3.66	-0.3005
Black, non-Hispanic	657	90.6	2.31	-0.5221
More than one race, non-Hispanic	76	97.9	1.23	4.4928
American Indian or Alaska Native, non-Hispanic	110	92.5	6.86	0.0988
White, non-Hispanic	6,671	92.4	0.63	0.6430
Region				
Northeast	1,465	86.1	1.72	-3.1513
Midwest	1,944	95.2	0.90	3.2268
South	2,976	93.2	0.92	1.2550
West	1,783	90.6	1.38	-0.8419
Teachers in sample of Hispanic ethnicity, any race				
Overall	395	82.8	3.27	-2.7216
Teacher's grade level				
Elementary	198	85.6	4.05	0.5335
Middle	86	87.1	6.54	0.5931
Secondary	111	71.2	10.17	-1.0852
School type				
Traditional public	239	82.9	3.79	0.0162
Public charter	35	93.3	5.22	1.6984
Private	121	79.5	6.15	-0.4710
Teaching experience				
New	175	85.1	4.29	0.4279
Experienced	220	81.9	4.45	-0.1673
Teacher status				
Leaver	53	95.8	2.53	3.1370
Mover	90	77.5	7.95	-0.6124
Stayer	127	82.4	4.07	-0.0844
Unknown	125	88.1	3.22	1.1631
Minority status				
Minority	392	82.7	3.28	-0.0170
Nonminority	3	98.0	31.26	0.4828

† Not applicable.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current Teacher Data File" and "Former Teacher Data File," 2004–05.

The frame distribution was compared to the base-weighted respondent distribution for the stratification characteristics (i.e., urbanicity, teacher's grade level, school type, teaching experience, teacher status, school size, percent minority enrollment, teacher's age, teacher's race/ethnicity, and region). Since the Hispanic response rate was below 85 percent, the distributions were compared for Hispanic teachers broken down by stratification characteristics. The results are as listed in table 11.

Table 11.   Base-weighted frame distribution of TFS teachers, interviewed distribution, and
**   *t* statistic, by selected reporting characteristics: 2004–05**

Reporting characteristic	Frame distribution (adjusted for out-of-scope teachers) and standard error		Interviewed distribution (adjusted for out-of-scope teachers) and standard error		*t* statistic (frame proportion compared to interviewed proportion)
	Proportion	Standard error	Proportion	Standard error	
Overall	1.0000	0.0137	1.0000	0.0237	0.0000
All teachers in sample					
Urbanicity					
Central city	0.2732	0.0046	0.2677	0.0107	-0.4759
Suburban	0.5038	0.0085	0.5040	0.0166	0.0139
Rural	0.2231	0.0111	0.2283	0.0128	0.3096
Teacher's grade level					
Elementary	0.4542	0.0172	0.4556	0.0226	0.0494
Middle	0.2494	0.0064	0.2512	0.0115	0.1378
Secondary	0.2964	0.0045	0.2932	0.0128	-0.2377
School type					
Traditional public	0.8780	0.0300	0.8788	0.0342	0.0172
Public charter	0.0108	0.0005	0.0109	0.0009	0.0883
Private	0.1112	0.0391	0.1103	0.0343	-0.0169
Teaching experience					
New	0.2236	0.0059	0.2218	0.0086	-0.1665
Experienced	0.7764	0.0111	0.7782	0.0212	0.0730
Teacher status					
Leaver	0.0277	0.0014	0.0281	0.0013	0.2169
Mover	0.0648	0.0032	0.0646	0.0034	-0.0274
Stayer	0.8499	0.0103	0.8513	0.0227	0.0599
Unknown	0.0577	0.0078	0.0559	0.0086	-0.1525
School size					
Less than 200	0.0895	0.0313	0.0827	0.0270	-0.1640
200–349	0.1363	0.0039	0.1347	0.0072	-0.1940
350–499	0.1796	0.0074	0.1916	0.0112	0.8962
500–749	0.2414	0.0088	0.2382	0.0139	-0.1916
750 or more	0.3532	0.0102	0.3527	0.0163	-0.0254
Percent minority enrollment					
Less than 10 percent	0.3117	0.0204	0.3321	0.0224	0.6734
10–34 percent	0.2850	0.0074	0.2796	0.0114	-0.3998
35 percent or more	0.4033	0.0111	0.3884	0.0170	-0.7338
Teacher's age					
Less than 30	0.1590	0.0049	0.1612	0.0087	0.2217
30–39	0.2453	0.0039	0.2385	0.0114	-0.5637
40–49	0.2420	0.0052	0.2527	0.0145	0.6935
50–64	0.3429	0.0115	0.3349	0.0138	-0.4424
65 or more	0.0108	0.0020	0.0127	0.0022	0.6196

See notes at end of table.

Table 11.   Base-weighted frame distribution of TFS teachers, interviewed distribution, and
***t* statistic, by selected reporting characteristics: 2004–05—Continued**

Reporting characteristic	Frame distribution (adjusted for out-of-scope teachers) and standard error		Interviewed distribution (adjusted for out-of-scope teachers) and standard error		*t* statistic (frame proportion compared to interviewed proportion)
	Proportion	Standard error	Proportion	Standard error	
All teachers in sample—Continued					
Teacher's race/ethnicity					
Hispanic, any race	0.0536	0.0033	0.0412	0.0042	-2.3260
Asian or Pacific Islander, non-Hispanic	0.0150	0.0007	0.0192	0.0031	1.3068
Black, non-Hispanic	0.0704	0.0028	0.0688	0.0054	-0.2563
More than one race, non-Hispanic	0.0075	0.0010	0.0099	0.0023	0.9808
American Indian or Alaska Native, non-Hispanic	0.0054	0.0008	0.0087	0.0020	1.5317
White, non-Hispanic	0.8480	0.0142	0.8521	0.0239	0.1448
Region					
Northeast	0.1942	0.0032	0.1839	0.0109	-0.9011
Midwest	0.2440	0.0047	0.2533	0.0114	0.7567
South	0.3802	0.0116	0.3843	0.0157	0.2097
West	0.1816	0.0041	0.1785	0.0091	-0.3147
Teachers in sample of Hispanic ethnicity, any race					
Overall	1.0000	0.0610	1.0000	0.1022	0.0000
Teacher's grade level					
Elementary	0.5544	0.0538	0.6241	0.0933	0.6471
Middle	0.1953	0.0226	0.1926	0.0441	-0.0548
Secondary	0.2503	0.0136	0.1833	0.0451	-1.4229
School type					
Traditional public	0.8942	0.0617	0.8871	0.1040	-0.0591
Public charter	0.0181	0.0025	0.0247	0.0053	1.1105
Private	0.0876	0.0155	0.0882	0.0156	0.0265
Teaching experience					
New	0.2814	0.0234	0.2939	0.0348	0.2986
Experienced	0.7186	0.0494	0.7061	0.0943	-0.1177
Teacher status					
Leaver	0.0124	0.0027	0.0169	0.0038	0.9636
Mover	0.0546	0.0077	0.0579	0.0092	0.2725
Stayer	0.8534	0.0566	0.8252	0.0980	-0.2492
Unknown	0.0795	0.0107	0.1000	0.0145	1.1355
Minority status					
Minority	0.9953	0.0604	0.9939	0.1016	-0.0118
Nonminority	0.0047	0.0027	0.0061	0.0054	0.2304

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current Teacher Data File" and "Former Teacher Data File," 2004–05.

Conclusion/Course of Action. Based on this analysis, evidence of substantial bias was not found.

Item Nonresponse Bias Analysis

Overview of Methodology

The item bias analysis examined the overall response rate for each item on both TFS data files. The analysis included examining the item response rates by the reporting characteristics (i.e., urbanicity, teacher's grade level, school type, school size, percent minority enrollment, teacher status, teaching experience, teacher's age, teacher's race/ethnicity, and region) using the final weight for all in-scope sampled units. If the overall response rate for the item fell below 70 percent, the item will be footnoted in NCES publications with "Item response rate is below 70 percent" as a method of cautioning the user that the low item response rate introduces some potential for bias in the imputation procedure. For any characteristic where the item response rate was less than 85 percent, a more detailed analysis was done by stratification characteristics. The results were highlighted if that particular cell had a significantly higher or lower response rate than the file as a whole and bolded if the difference was noteworthy. A noteworthy difference met the following conditions:

- The difference relative to the overall response rate for the particular item was greater than 10 percent.
- The absolute difference was greater than one percentage point.
- The coefficient of variation was less than 15 percent.
- The cell had at least 30 interviews.

Table 12 presents the number of items by response rate for each file. Of particular concern are the items with an overall response rate below 70 percent. These items are listed in exhibit 2.

Table 12.   Number of questionnaire items, by response rate category and data file: 2004–05

	Total	Number of items with a response rate of…			
Data file	number of items	95.0 percent or above	85.0 to 94.9 percent	70.0 to 84.9 percent	Below 70 percent
Current teacher	170	126	37	6	1
Former teacher	100	41	53	2	4

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current Teacher Data File" and "Former Teacher Data File," 2004–05.

Exhibit 2.   Items with a response rate below 70 percent, by data file: 2004–05

Data file	Item
Current teacher	Item 44: How many family members living in your household during 2004 were under the age of 5?
Former teacher	Item 26: Would any factors other than the ones listed above influence your decision to return to the position of a K–12 teacher? Item 26 factors: If "Yes," what factors? Item 31: How many family members living in your household during 2004 were under the age of 5? Item 35 (other): Where is your access to the Internet—at home, work, or elsewhere?

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current Teacher Data File" and "Former Teacher Data File," 2004–05.

Summary of Conclusions.

Current Teachers—Seven items had a response rate below 85 percent, requiring a closer examination. One item had a response rate below 70 percent, necessitating a footnote. A closer examination of response rates revealed no substantial evidence of a bias.

Former Teachers—Six items had a response rate below 85 percent, requiring a closer examination. Four items had a response rate below 70 percent, necessitating a footnote. A closer examination of response rates revealed no substantial evidence of a bias.

Chapter 6. Data Processing

Once the 2004–05 Teacher Follow-up Survey (TFS) data collection was completed, data processing began. U.S. Census Bureau clerical processing staff in Jeffersonville, Indiana, was responsible for the first phase of data processing. This involved using the Automated Tracking and Control (ATAC) system to assign a check-in code for each paper TFS questionnaire received to indicate its status; TFS questionnaires completed on the Internet were automatically assigned check-in codes by the internet instrument. The data from completed questionnaires then were captured (converted from paper to electronic format) and sent to Census Bureau analysts in reformatted SAS datasets so the analysts could more easily review the data. Census Bureau analysts were responsible for resolving outcome codes, conducting preliminary data reviews, and assigning preliminary interview status. They performed a series of computer edits on the data to identify and resolve inconsistencies and assigned a final interview status to each case. The final interview status is the final determination of whether each case was an interview, a noninterview, or out of scope for TFS, using stricter criteria than those used to determine the preliminary interview status of each TFS case. After these steps were completed, there were still items that were "not answered." These items were imputed during the final stages of data processing. Once all stages of data processing were completed, the final data products were created.

Questionnaire Check-in

Check-in of Paper Questionnaires

The Census Bureau clerical processing staff received questionnaires directly from both the sample teachers and from the 12 Regional Offices' field follow-ups.[16] Upon receipt, staff assigned a check-in code (e.g., completed questionnaire, blank questionnaire, refusal, teacher deceased) to each questionnaire to indicate its status. A paper questionnaire was checked-in as a "completed questionnaire"[17] when the respondent completed at least one item. The exception to this rule was when item 1 on either questionnaire indicated that the respondent was sent the incorrect questionnaire type due to the former school's inaccurate reporting of the respondent's teaching or other occupational status on the Teacher Status Form[18]; these respondents were assigned a unique "switcher" check-in code indicating this. See chapter 4 for further information about "switchers." The remaining check-in codes were assigned based upon any notes or indicators written on the cover of or attached to the returned questionnaire. If staff members were unsure of what check-in code to assign, they sent the case to Census Bureau headquarters analysts for reconciliation.

All TFS questionnaires were assigned a check-in code. The code was entered into the ATAC system. If there was a change to the address either marked on the questionnaire label or indicated by the post office, the address information was updated in the ATAC system as well. The questionnaires were then grouped

---

[16] The Census Bureau has Regional Offices in Atlanta, Boston, Charlotte, Chicago, Dallas, Denver, Detroit, Kansas City, Los Angeles, New York, Philadelphia, and Seattle.

[17] The check-in code indicating a "completed questionnaire" does not necessarily indicate that a case is a "complete interview." Interview status was assigned both during the preliminary and final interview status (ISR) stages of data processing. See the "Preliminary ISR Classification" and "Final Interview Status Edit" sections of this chapter for a detailed description of the criteria for former and current teacher questionnaire complete interviews.

[18] The Teacher Status Form was mailed to TFS sampled schools that provided lists of teachers during the 2003–04 Schools and Staffing Survey. These schools were asked to complete the form by indicating whether each teacher listed was still teaching in that school (stayer), was teaching in another school (mover), or left the teaching profession all together (leaver).

into batches by questionnaire type, doc type,[19] and check-in code. Only completed interviews were sent on for the next step of data processing, which was data capture.

Check-in of Internet Questionnaires

Approximately 12 percent of respondents completed the internet version of the TFS questionnaire that included the questions for both the current and former teacher questionnaires. As respondents submitted their questionnaires online, the internet instrument automatically edited them for critical items, which are those items that must be answered in order for a questionnaire to be considered complete. While the paper questionnaires were assigned a check-in code of "completed questionnaire" if at least one item was completed, the internet questionnaires were assigned a check-in code of complete as long as long as all critical items were complete. The automated edits on these internet forms helped to reduce the amount of editing and review that these cases would have had to go through at later stages of data processing. For a description of the critical items for TFS, refer to exhibit 3 in the "Preliminary Data Review" section of this chapter.

Several of the critical items on the TFS internet version served as screener questions because the respondent's answers to these questions determined which questionnaire path the respondent would follow. The combination of the first two items determined whether the respondent went down the leaver (i.e., former teacher) path or stayer/mover (i.e., current teacher) path. The first item asked the respondent if he or she currently taught any regularly scheduled classes in any of grades prekindergarten through 12. If not, the respondent would automatically follow the former teacher path of questions. Otherwise, the second item asked the respondent to classify his or her current position at the current school.

As long as the respondent was not a short-term substitute, a student teacher, or a teacher aide, he or she would follow the current teacher path of questions.[20] Further down the current teacher questionnaire path, a question asked whether the respondent was teaching at the same school as previous year. This question determined whether the current teacher respondent would follow the "mover" path of questions.

If all of the above screener items (when applicable, depending on the questionnaire type) as well as a few additional items were completed, the check-in code automatically reflected that the case was a complete interview. See the "Final Interview Status Edit" section of this chapter for a detailed description of the criteria for former and current teacher questionnaire complete interviews.

Data Capture

Data Capture of Paper Questionnaires

The 2004–05 TFS data were captured (i.e., converted from paper to electronic format) using manual data keying. During check-in, the questionnaires were split into groups called "batches" by questionnaire type, doc type, and check-in code, and then they were manually keyed. Manual data keying for both TFS questionnaires was accomplished using a Key from Paper (KFP) data capture system. Analysts wrote specifications for data keying, and programmers used these specifications to develop the KFP system for each survey prior to keying. It was programmed to present screens of questionnaire items to data keying

---

[19] The doc type indicates if the questionnaire that was received by the clerical staff was from the first mailing, second mailing, or field follow-up.

[20] For TFS, teacher aides, student teachers, and short-term substitute teachers were not considered regular classroom teachers. If a respondent was a regular classroom teacher during the 2003–04 Schools and Staffing Survey school year and changed assignments to one of these three positions for the 2004–05 school year, then he or she was considered to be a leaver and should have completed the former teacher questionnaire.

staff, who worked through each page of the questionnaire and keyed any entries into the appropriate fields on the screens.

All KFP entries were 100 percent verified by the keying staff, meaning that each field was keyed twice and the results were compared automatically for discrepancies and, subsequently, verified. The verification during this operation allowed up to a 1 percent error on a field-to-field basis. Error rates were calculated by dividing the total number of keying errors by the total number of keyed fields.

If an entire batch of questionnaires had a total error of more than 1 percent (i.e., all keying errors for that batch divided by the total keyed fields in that batch exceeded 1 percent), the batch was unacceptable, and all questionnaires within the batch were 100 percent verified a second time. A more detailed discussion of data capture and results of the keying verification for the TFS questionnaires are provided in "Appendix F. Quality Assurance for TFS Keying and Mailout Operations."

Images of each questionnaire were captured after data entry was completed. The image files were used during subsequent steps of data processing to view the images of the actual questionnaires electronically.

Internet Questionnaires

Data for the internet questionnaires did not go through a separate data capture operation; as respondents completed questions on the TFS website, data were automatically captured and saved by the system. At this point, the data were already in electronic format. Unlike the TFS paper questionnaires, there were no images of the internet questionnaires to be captured and stored as image files (since this was a web-based survey). Therefore, during subsequent steps of data processing, Census Bureau analysts were not able to refer back to the actual questionnaires as a reference.

Reformatting

After the paper questionnaire data were captured, the output files were reformatted into SAS datasets. The internet data were in a different electronic format and also needed to be reformatted. Census Bureau analysts provided specifications to programmers that mapped the internet variables to the variables in the reformatted paper questionnaire data. This allowed analysts to proceed with data processing and cleaning with the paper and internet data together in the same SAS datasets.

Preliminary Data Review

Once all of the data were reformatted, Census Bureau analysts began an extensive preliminary data review process. During data review, analysts examined frequencies of each data item in order to identify any suspicious values (e.g., if an item's response was an extreme value or outside of the range of possible answer choices or if an answer seemed unlikely given the respondent's other responses within the survey). When suspicious values were found, analysts looked at the image of the questionnaire (for paper questionnaires only) to verify that the data were keyed correctly. Fixes were made to the data files when necessary.

The TFS data were split into two different files when undergoing the preliminary stages of data review. There was a former teacher data file for leavers and a current teacher data file for stayers and movers. During data review, analysts checked that the appropriate cases were in each data file. They also checked survey counts by type to make sure that there was consistency between the actual counts of stayers, movers, and leavers and the variable that programmers created to indicate stayer/mover/leaver status. This

variable kept track of "switcher cases" (i.e., cases where the respondent's actual teaching or other occupational status differed from that reported by the school on the Teacher Status Form).

In addition, analysts reviewed cases in which any critical item was not answered or an insufficient number of other questionnaire items were completed. Analysts viewed these cases as potential problems and verified that data were keyed correctly by reviewing an electronic image of the questionnaire (for paper questionnaires only). See exhibit 3 for a detailed list of critical items for TFS.

Exhibit 3.   TFS critical item table, by questionnaire: 2004–05

Questionnaire	Page	Item	Source code[1]	Description
Questionnaire for Former Teachers— 3 critical items	5	1a	F0050	Do you currently teach any regularly scheduled class(es) in any of grades pre-K–12? (Response should be "No" in order to continue with the former teacher questionnaire path.)
	OR:			
	5	1c	F0051	How do you classify your position at your current school, that is, the activity at which you spend most of your time during the school year? (Response should be short-term substitute, student teacher, or teacher aide in order to continue with the former teacher questionnaire path.)
	6	2	F0553	What is your main occupational status?
Questionnaire for Current Teachers— 4 critical items	5	1a	F0050	Do you currently teach any regularly scheduled class(es) in any of grades pre-K–12? (Response should be "Yes" in order to continue with the current teacher questionnaire path.)
	5	1b	F0051	How do you classify your position at your current school, that is, the activity at which you spend most of your time during the school year? (Response should **not** be short-term substitute, student teacher, or teacher aide.)
	8	7	F0058–F0072	In which grades are the students you teach at your current school?
	14	19a	F0136	Are you currently teaching in the same school as you were last year (2003–04)?[2]

[1] Source codes are used to identify specific items on the Schools and Staffing Survey questionnaires. For each questionnaire item, the four-digit source code can be found to the left of the first answer choice.
[2] Although each respondent's stayer or mover status was provided by his or her 2003–04 school on the Teacher Status Form, it was not always accurate. This item determined each respondent's final stayer/mover status. If he or she responded "yes" to the question, "Are you currently teaching in the same school as you were last year (2003–04)?" then the respondent was a stayer. If the respondent responded "no," then he or she was a mover.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

If critical item data were in fact missing, analysts attempted to use nonintrusive means of obtaining the data (e.g., imputation from other items on the questionnaire) so that these cases would not become noninterviews due to insufficient data in later stages of data processing.

Finally, Census Bureau analysts were responsible for resolving the ATAC check-in codes of a handful of former and current teacher questionnaire cases during the preliminary stages of data review. This occurred in cases in which both former and current teacher questionnaires was received by the clerical processing staff. In some cases, both were completed interviews, and analysts had to determine which

questionnaire represented the respondent's current situation by looking at images of both questionnaires. The analysts adjusted the check-in codes to indicate which to keep as a complete interview and which to drop.

In other cases, more than one of the same questionnaire type was received from a particular respondent, and analysts had to remove the duplicate case by deciding which questionnaire to keep and which to drop based upon the data in each independent questionnaire.

Preliminary Interview Status Recode (ISR) Classification

The next step in data processing was the preliminary determination of each case's interview status recode (ISR); that is, whether each case was an interview, a noninterview, or out of scope for TFS. In general, cases with an "out-of-scope" outcome code (ISR = 3) for the preliminary ISR were teachers who were deceased, had moved out of the United States, or had never been teachers (e.g., school administrators who had reported incorrectly on the Schools and Staffing Survey [SASS] during the previous school year). An out-of-scope code for each out-of-scope case was created to indicate the reason why the case was out of scope.

Cases with data entries were classified as completed interviews (ISR = 1). Cases without any data or cases in which the respondent refused to complete the survey were classified as noninterviews (ISR = 2). A noninterview code was created for each noninterview case to indicate the reason why the case was a noninterview.

During the data review of the preliminary ISR data, Census Bureau analysts looked at the specific cases that were checked-in as "complete questionnaires" but changed to out of scopes or noninterviews during the preliminary ISR classification. Analysts looked at the electronic images of these cases and, in some cases, found notes and other indications from the respondent within the questionnaire that led the analysts to believe that an inconsistent out-of-scope or noninterview code had been assigned. They reassigned appropriate codes to these cases during this stage of data processing.

A more detailed discussion of interview status can be found in the "Final Interview Status Edit" section of this chapter.

Computer Edits

After preliminary data review and preliminary ISR classification, the cases that were classified as complete interviews during the preliminary ISR were submitted to a series of computer edits. These edits consisted of a range check, a consistency edit, and a blanking edit.

The first of the computer edits was the range check, which was used to delete entries that were outside the range of acceptable values that were set prior to the administration of TFS. For example, the range of acceptable values for item 38 on the current teacher questionnaire ("During the current school year, what is your academic year base teaching salary?") was $0 to $200,000. Therefore, a salary entry of $310,000 would be deleted and recoded as a missing item by the range check. The purpose of this check was to eliminate any extreme values, such as this $310,000 salary, a highly unlikely base salary for a teacher.

Actual changes to the data were made during the consistency edit. The consistency edits identified inconsistent entries within each case and, whenever possible, corrected them. If the inconsistencies could not be corrected, the entries were deleted and recoded as missing items. These inconsistencies were

- within items (e.g., if the yes/no part of current teacher questionnaire item 37a—whether the respondent had any earnings from teaching summer school in the current or any other school— was "no," but the earning amount was greater than zero for the second part of the item); and
- between items (e.g., if former teacher questionnaire item 4c indicated that the respondent was working without pay in a family business, farm, or volunteer job, but item 6 indicated that the earnings at this job were greater than zero).

In addition, the consistency edit filled in some items where data were missing or incomplete by using other information on the same data record. For example, if former teacher questionnaire item 9 ("Do you consider yourself to be retired from the position of a K–12 teacher?") was not answered, but item 2 ("What is your main occupational status?") was marked "retired," then "yes" was marked for item 9 during the consistency edit.

The blanking edits deleted extraneous entries (e.g., in situations where skip patterns were not followed correctly) and assigned the "not answered" code to items that the respondent should have answered but did not.

The only records that were put through this series of edits were those classified as completed interviews during the preliminary ISR. The tables in "Appendix G. Changes Made to TFS Variables During Computer Edits, by Data File" show the number of edit changes made to entries for each of the variables within each data file. These changes are summarized in table 13 below.

Table 13.   Summary of changes made to variables in the computer edits, by data file: 2004–05

Data file	Total number of cases	Total number of variables in questionnaire	Number of variables changed during edits, by percentage of all TFS data records on which the variable was changed			
			None	1–15 percent	16–30 percent	More than 30 percent
Former teacher	2,681	104	34	69	1	0
Current teacher	4,806	200	121	79	0	0

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

Final Interview Status Edit

After the range checks, consistency edits, and blanking edits were run and reviewed by analysts, the records were put through an edit to make a final determination of whether the case was eligible for the survey and, if so, whether sufficient data had been collected for the case to be classified as a completed interview. A final ISR value was assigned to each case as a result of this edit.

1. **Questionnaire for Former Teachers**
   - A case was classified as **out of scope** (ISR = 3) if
     - the sampled person was deceased; or
     - the sampled person moved outside of the United States; or
     - the sampled person had never been a teacher.
   - A case was classified as an **interview** (ISR = 1) if
     - none of the conditions for out-of-scope cases was met; and
     - the sampled person reported that he/she did not currently teach any regularly scheduled classes in any of grades pre-K–12 (F0050) or reported that his/her job classification was a short-term substitute, student teacher, or teacher aide (F0051); and

o the sampled person reported his or her main occupational status (F0553); and
o there were data in at least 10 percent of the remaining items (nine answered items).
- A case was classified as a **noninterview** (ISR = 2) if an eligible case did not meet the requirements to be an interview case.

2. **Questionnaire for Current Teachers**
   - A case was classified as **out of scope** (ISR = 3) if
     o the sampled teacher was deceased; or
     o the sampled teacher had moved outside of the United States; or
     o the sampled person had never been a teacher.
   - A case was classified as an **interview** (ISR = 1) if
     o none of the conditions for out-of-scope cases was met; and
     o the sampled teacher indicated that he/she taught any regularly scheduled class(es) in any of grades pre-K–12 (F0050); and
     o the sampled teacher reported his/her job classification (F0051); and
     o the sampled teacher reported the grades in which the students that he/she currently taught were (F0058–F0072); and
     o the sampled teacher indicated whether he/she was teaching in the same school as the previous year (F0136); and
     o there were data in at least 10 percent of the remaining items (16 answered items).
   - A case was classified as a **noninterview** (ISR = 2) if an eligible case did not meet the requirements to be an interview case.

The preliminary ISR and final ISR counts for each data file and the percentage of change for each ISR classification are shown in table 14.

Table 14. Preliminary and final interview status recode (ISR) counts and percentage change, by ISR category and data file: 2004–05

Data file	Sample size	Preliminary ISR			Final ISR			Percentage change[1]		
		Inter-views	Non-inter-views	Out of scope	Inter-views	Non-inter-views	Out of scope	Inter-views	Non-inter-views	Out of scope
Former teacher	3,039	2,681	300	58	2,653	328	58	-1.04	9.33	0
Current teacher	5,218	4,806	381	31	4,776	411	31	-0.62	7.87	0

[1] The percentage change for each ISR category (interview, noninterview, out of scope) is computed by subtracting the number of cases in the preliminary ISR count from the number of cases in the final ISR count, and dividing by the number of cases in the preliminary ISR count.
NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

Creating Imputation Flags

After the final ISR edits were run, there were still cases with "not answered" values on the files for some variables. Values were created for these items in the next step of the processing, which was imputation. There were three different stages of imputation: stage 1, stage 2, and clerical imputation. After each stage of imputation, the computer edits (described in the "Computer Edits" section of this chapter) were re-run to ensure that the imputed data were consistent with the existing questionnaire data. The imputation stage of data processing is described further in "Chapter 7. Imputation Procedures."

Because the consistency edits and imputations made actual changes to the TFS data, different processing flags were created to indicate these changes. Flags that were used in the consistency edit stage of data processing are in the format ef_[source code]+1. That is, each time a change was made to an item due to a consistency edit, the value of the flag for that item increased by one. This way analysts could keep track of how many times a consistency edit changed the data throughout the duration of data processing.

Flags that were used in the imputation stage of data processing were different than those used for consistency edits. Rather than simply counting how many times an imputation edit changed data, imputations flags were assigned specific values to indicate exactly how items were being imputed. Imputation flags are in the format f_[source code] = __. The definitions for each imputation flag used in the TFS are described in exhibit 4 below.

Exhibit 4.   Flags used in processing TFS questionnaires: 2004–05

Imputation flag values	Description of flag
f_[source code] = 0	Not imputed
f_[source code] = 1	Original value was ratio adjusted
f_[source code] = 2	Value was imputed by using data from other variables in same record
f_[source code] = 3	Value was imputed by using data from school or teacher record
f_[source code] = 7	Imputed from donor
f_[source code] = 8	Clerical imputation

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

Data Products

After all stages of imputation were completed and the blanking and consistency edits were run once again, the data were split into two files by questionnaire type. One file contained the records for current teachers (Questionnaire for Current Teachers data) and the other contained records for former teachers (Questionnaire for Former Teachers data).

Each of these files included all variables, including frame variables, survey variables, created variables, weighting variables, and imputation flags. These files were used as the source files for the restricted-use datasets, in which some of the processing and sampling variables were removed, and data nondisclosure procedures were implemented. (See chapter 10 for additional description of the restricted-use data files.)

Chapter 7. Imputation Procedures

Overview

Once the computer edits were run on the 2004–05 Teacher Follow-up Survey (TFS) data and reviewed by U.S. Census Bureau analysts, "not answered"[21] items still remained in the datasets. (For more details on computer edits see chapter 6.) In order to fill these items with data, questionnaires were put through three independent stages of imputation. From one stage to the next, more assumptions were made about how the participant might have responded to the item. The first stage of imputation involved using items from the same TFS questionnaire or items from the corresponding 2003–04 Schools and Staffing Survey (SASS) school or teacher questionnaire to impute data for the missing items. Any remaining unanswered items were imputed in the second stage of imputation. This stage included "hot deck" imputation (in which donor records were established and used to impute data), ratio imputation by subsamples of data, and random ratio imputation by subsamples of data. Once the first two stages of imputation were completed, any remaining unanswered items were imputed clerically by Census Bureau analysts in the third and final stage of imputation.

Each time an unanswered TFS item was imputed, a numerical flag corresponding to the stage and type of imputation was assigned to the imputed item. These flags identify to data users which items were imputed, and the method by which they were imputed. Therefore, data users can decide whether to include imputed data in their analyses, and if so, which types of imputed data they want to use.

First-Stage Imputation

During the first stage of imputation, unanswered survey items were imputed with a response using data either from other items in the same TFS questionnaire or from items in the related 2003–04 SASS school or teacher questionnaires. Imputed values were valid responses because they were within the range of acceptable values that was set prior to the administration of TFS and were consistent with the respondent's answers to related items. During this stage of imputation, data from the Questionnaire for Former Teachers and the Questionnaire for Current Teachers were kept in two separate files.

- *Imputing with other items on the same TFS questionnaire record*. Based on entries from related items on the TFS record, assumptions were made about how the respondent might have answered items. For example, item 8 on the current teacher questionnaire asked how the sampled teacher's classes were organized. If the teacher indicated that he or she was a general kindergarten or elementary grade teacher in item 5 (main teaching assignment field), then item 8 was marked to indicate that the respondent instructed the same group of students all or most of the day in multiple subjects (self-contained class).
- *Imputing with related items on the 2003–04 SASS school or teacher record*. Since each TFS sampled teacher participated in the 2003–04 SASS, information from the SASS record was used to impute values in the first stage of imputation. For example, item 17 on the former teacher questionnaire (item 28 on the current teacher questionnaire) asked if any of the respondent's students participated in a required state or district assessment program in the subject that the respondent taught the prior school year. If this item was blank and the respondent indicated in the 2003–04 SASS teacher questionnaire that his or her students did, in fact, participate in such an assessment program, then the item was marked "yes"; otherwise, it was marked "no."

---

[21] Some items were correctly left blank by the respondent because of skip patterns in the questionnaire. For these respondents, the items were changed to "valid skips" during the blanking edit stage of data processing. Valid skips were not imputed and were not classified as "not answered" items.

In addition, many sampled teachers' corresponding school completed the 2003–04 school questionnaire; if this was the case, information from the SASS school record was used to impute values during the first stage of imputation.

Since there were two different sources for stage 1 imputation, each source was assigned a separate value for the stage 1 imputation flag. The definitions of these flags are as follows:

0   Data were reported by the sampled teacher. No adjustment or imputation was made.
2   The item was imputed based on data from another item within the same TFS questionnaire.
3   The item was imputed based on data from the 2003–04 SASS school or teacher questionnaire associated with the TFS sampled teacher.

Tables 15 and 16 at the end of the chapter include a summary of the amount of imputation performed in stage 1 processing.

Second-Stage Imputation

Those items still unanswered after stage 1 imputation entered the second stage of imputation. Before entering this stage, the former and current teacher questionnaire datasets were split up further. The former teacher questionnaire data file became two separate files; one file contained former teacher questionnaire data for sampled teachers from traditional public and public charter schools, and the other file contained former teacher questionnaire data for sampled teachers from private schools.

Two main approaches were used in the second stage of imputation for TFS. In one approach (hot deck imputation), data were imputed from items found on questionnaires of the same type (former or current teacher questionnaire) that had certain characteristics in common. These records are called "donor records." In the second approach to stage 2 imputation, data were split into subsamples based upon grade levels offered in their 2003–04 school (in subsample ratio imputation) or other variables related to the item being imputed (in random subsample ratio imputation), and items were imputed so that the distribution of responses in each of these subsamples was preserved. Both hot deck imputation and subsample ratio imputation are explained further in this chapter.

Items that were imputed during the second stage of imputation were assigned a particular numerical flag. The definition of the flag is as follows:

7   The item was imputed during the second stage of imputation by hot deck or subsample ratio imputation.

Hot Deck Imputation

During hot deck imputation, donor records were established, and imputation was based on the data within the donor's record. Donors were selected based on the way they answered specific items; these items are called "matching variables." If two sampled teachers answered the selected matching items the same way, it was assumed that they were comparable and that imputation of one data item from the donor record was reasonable and would provide a response similar to one the sampled teacher would have given on his or her own.

The matching variables used to establish donor relationships were selected based on the type of data the donor would supply to the record being imputed. For example, if a current teacher questionnaire respondent's answer to a given item might be influenced by the respondent's teaching field and the

proximity of the school to a city, then these variables were used to find another sampled teacher in a school with similar characteristics.

The datasets were sorted by matching variables in the order of their importance, that is, how much influence the matching variables had in predicting the value of the variable being imputed. Sorting the data helped to ensure that the appropriate donors were the most similar to the record undergoing imputation.

For stage 2 hot deck imputation among ***public school sampled teachers***, the states in which the sampled teachers' 2003–04 schools (as reported in SASS) were located were combined into four groups by their geographic region in order to increase the size of the data pool. These four regions were the Northeast, Midwest, South, and West. All imputation was done within the geographic region group; that is, the donor record had to be from a sampled teacher within the same region as the incomplete record.

- *Former public school sampled teachers*. Within each region group, the records were sorted by the following variables:
  - o STATE   State in which respondent's 2003–04 school was located;
  - o WORK   Whether the respondent was currently working; and
  - o T0416   Sampled teacher's birth year (from which age was derived).

  The records were sorted by STATE / WORK / T0416.

- *Current public school sampled teachers*. Within each region group, the records were sorted by the following variables:
  - o STATE   State in which respondent's 2003–04 school was located,
  - o STAYER   Whether the respondent was teaching at the same school as during the 2003–04 school year,
  - o TFSLEVEL   Grade levels taught as reported in TFS, and
  - o TEACHCODE   Respondent's main teaching assignment field during the 2003–04 school year.

  The records were sorted by STATE / STAYER / TFSLEVEL / TEACHCODE.

For stage 2 hot deck imputation among ***private school sampled teachers***, the typologies of the sampled teachers' 2003–04 schools (as reported in SASS) were combined into three groups of affiliations. These three affiliations were Catholic, other religious, and nonsectarian. All imputation was done within the affiliation group; that is, the donor record had to be from a sampled teacher within the same affiliation group as the incomplete record.

- *Former private school sampled teachers*. Within each affiliation group, the records were sorted by the following variables:
  - o TYPOLOGY   Typology of respondent's 2003–04 school;
  - o WORK   Whether the respondent was currently working; and
  - o T0416   Sampled teacher's birth year (from which age was derived).

  The records were sorted by TYPOLOGY / WORK / T0416.

- *Current private school sampled teachers*. Within each affiliation group, the records were sorted by the following variables:
  - o TYPOLOGY   Typology of respondent's 2003–04 school;
  - o STAYER   Whether the respondent was teaching at the same school as during the 2003–04 school year;
  - o TFSLEVEL   Grade levels taught as reported in TFS; and

    o   TEACHCODE   Respondent's main teaching assignment field during the 2003–04 school year.

The records were sorted by TYPOLOGY / STAYER / TFSLEVEL / TEACHCODE.

When there were not enough donor records within any given stratification cell, a collapsing routine was instituted. This was done to ensure that values inconsistent with other data on the same record would not be imputed simply because a record was close to the boundary between the stratification cells (e.g., there were other records that were suitable donors or the record was not similar enough to be a donor). For example, for the current teacher questionnaire, the collapsing routine for the matching variable TEAEXPER (the sampled teacher's level of teaching experience) is shown in the matrix below. Each row of the matrix represents a separate situation. An explanation follows.

> (1, 2, 0,
> 2, 3, 1,
> 3, 2, 0)

The values for the variable TEAEXPER are as follows:

- TEAEXPER = 0   No donor record was found for the case.
- TEAEXPER = 1   The sampled teacher has been teaching full and/or part time for 3 years or fewer.
- TEAEXPER = 2   The sampled teacher has been teaching full and/or part time for 15 years or fewer.
- TEAEXPER = 3   The sampled teacher has been teaching full and/or part time for 16 or more years.

If the value for TEAEXPER on the record with missing data was 1 and there was no available donor where TEAEXPER = 1, the collapsing program looked for a donor where TEAEXPER = 2. It did not look for cases where TEAEXPER = 3. When the collapsing routine hit 0, there was no donor available for this case. Likewise, if the value for TEAEXPER on the record with missing data was 2 and there was no available donor where TEAEXPER = 2, the collapsing program searched for a donor where TEAEXPER = 3, then TEAEXPER = 1. Finally, if the value for TEAEXPER on the record with missing data was 3 and there was no available donor where TEAEXPER = 3, the collapsing program looked for a donor where TEAEXPER = 2. It did not look for cases where TEAEXPER = 1. When the collapsing routine hit 0, there was no donor available for this case. In instances where no donor was available, the value was either imputed using subsample ratio imputation or was clerically imputed in the third stage of imputation.

Once the donor relationship was established, the donor record provided data items either directly or indirectly to the imputed record. For some items, such as item 8 on the former teacher questionnaire (how long the respondent planned to remain in his or her current position), data were copied directly to the record with the missing data. For other items, entries on the donor records were used as factors along with other data from the questionnaire to fill the incomplete item. For example, the unanswered item requesting the sampled teacher's current salary (Questionnaire for Current Teachers item 38) was filled by accepting the donor's ratio of the previous year's salary (T0399 on the SASS teacher questionnaires) to the current year's salary and applying the ratio to the sampled teacher's previous year's salary on the imputed record.

If an unanswered item was imputed using a donor record and the donor answered using the "other" option, then the write-in "please specify" portion was typically not imputed. In addition, open-ended

items that required the sampled teacher to write down his or her response, rather than selecting a provided response, were not imputed. This is because many of the write-in items asked about things that were very specific to each sampled teacher. For instance, Questionnaire for Former Teachers item 4a asked for the usual activities or duties at the respondent's job. Items such as these were not imputed and were given a value of -9, meaning "missing data," in the final data files.

There was one exception to this for an item that was not very respondent-specific; this exception was Questionnaire for Former Teachers item 35/Questionnaire for Current Teachers item 48 which asked where the respondent had access to the Internet. If a donor answered using the "other" option, the write-in "where" portion was also imputed.

In order to prevent a single record from having an undue impact on the data, a record could only be used as a donor a maximum of five times. There were no exceptions to this procedure.

Subsample Imputation

After the hot deck portion of the stage 2 imputation was completed, the data for sampled teachers from traditional public and public charter schools were merged with the data for sampled teachers from private schools by questionnaire type (i.e., former or current teacher questionnaire). Therefore, all former teacher questionnaire data were in one file, and all current teacher questionnaire data were in a second file. Data remained in these two files for the subsample ratio and random subsample ratio portions of the stage 2 imputations and the stage 3 clerical imputations.

Subsample Ratio Imputation

Items that remained unanswered after the hot deck imputation and that were categorical variables were imputed using subsample ratio imputation. First, data were divided into three subsamples (or groupings) based on the grade levels offered at the sampled teacher's school from the 2003–04 SASS school year (LEVEL). The groups were defined as follows:

- LEVEL = 1   Grades K through 6 (elementary);
- LEVEL = 2   Grades 7 through 12 (secondary); and
- LEVEL = 3   Combined grades (K through 12 and ungraded).

Next, the distribution (ratio) of responses to each unanswered categorical item was found within each LEVEL grouping (subsample). Finally, the unanswered items were assigned answers according to the subsample to which they belonged in order to preserve the response ratios within that subsample.

For example, among the subsample of LEVEL = 1 schools where a particular item with four answer categories was answered, that item had the following distribution (ratios) of answers: 40 percent chose the first category, 20 percent chose the second, 30 percent chose the third, and 10 percent chose the fourth category. This distribution would then be used to impute the records with missing data for this item where LEVEL = 1, so that the distribution of responses was preserved. Similarly, the percentage distribution of responses for records where LEVEL = 2 was calculated and the records with missing data that had the same LEVEL value were imputed accordingly. This operation was performed for all LEVEL values (subsamples).

Random Subsample Ratio Imputation

Random subsample ratio imputation is a method similar to subsample ratio imputation. However, it is more effective at handling items that are continuous variables—numeric variables for which the respondent is not provided with a fixed set of answer choices (e.g., salary). The program sorted the data into groupings based on the value of any variables that might have impacted the sampled teacher's answer to the item with missing data. Continuous variables were assigned a random probable value (e.g., a value between the 5th and 95th percentile) to cases with missing responses based on the range of values provided by sampled teachers with similar characteristics.

Census Bureau analysts determined which variables might have impacted the sampled teacher's answer to the item with missing data and used them to define a subsample within the dataset. For example, when imputing Questionnaire for Former Teachers item 27, the lowest acceptable teaching salary that would be required for the sampled teacher to return to teaching, the data were sorted into subsets based on

- item 25f, how an increase in salary would influence the decision to return to teaching;
- item 13f, the influence salary had on the respondent's decision to leave teaching;
- item 29, the respondent's total combined family income; and
- the base salary received as a teacher the prior year as indicated on the 2003–04 SASS survey.

If a record within this subsample had an unanswered item, an answer was randomly assigned so that the distribution of responses to that item remained the same for all records within the subsample. For example, if teachers who indicated that an increase in salary would be "extremely important" in influencing the decision to return to teaching reported

- that better salary and benefits were "very important" in the decision to leave,
- a combined family income of less than $35,000,
- a previous base teaching salary of less than $20,000, and
- a value for item 27 that ranged from $20,000 in the 5th percentile to $50,000 in the 95th percentile,

then the imputation procedure randomly assigned values to teachers with missing data that fell within this percentile range.

Tables 15 and 16 at the end of the chapter include a summary of the amount of imputation performed in stage 2 processing.

Third-Stage Imputation

In some cases, items still remained "not answered" after the first two stages of imputation. This happened when there was no available donor or the value imputed by the computer was inconsistent with values in other items.[22] Historically, certain types of items on SASS and TFS questionnaires have been imputed clerically during a third stage of imputation. For the 2004–05 TFS, all items still unanswered after the first two stages of imputation were imputed clerically during this third stage of imputation.

Items that were imputed during the third stage of imputation were assigned a particular numerical flag, which was defined as follows:

---

[22] After the second stage of imputation was complete, the blanking and consistency computer edits were re-run. If an imputed entry was inconsistent, it was adjusted or deleted by the consistency edit. For more information about computer edits, see chapter 6.

8   The item was imputed clerically during the third stage of imputation.

In order to determine an appropriate value for each unanswered item, Census Bureau analysts reviewed

- the original image of the questionnaire (for paper questionnaires only) to see if the sampled teacher had made any notes in the margin that might provide insight;
- other items within the same record with related information;
- similar cases to get an understanding of what the sampled teacher might have answered; and
- average responses for that item.

Tables 15 and 16 include a summary of the amount of imputation performed in the third stage of processing.

Post-Imputation Processing

After both the second and third stages of imputation were completed, the computer edits were re-run and any remaining data issues were resolved by Census Bureau analysts. (See chapter 6 for details on computer edits.) At this point, any item that was imputed more than 15 percent of the time was analyzed as part of the item bias analysis. (See chapter 5 for details about nonresponse bias analysis.) The computer edits were run again to ensure that values imputed in each stage of imputation were within acceptable ranges and were consistent with other items on the questionnaire.

Final File Imputation Tables

After the hot deck portion of stage 2 processing, the datasets were merged so that the data were in two files: one file for the former teacher questionnaire data and the other for the current teacher questionnaire data. The number of source codes[23] (specific items) that were imputed for a given percentage of records during a given stage of processing appears for each file below in tables 15 and 16. For example, during stage 1 imputation 12 survey items were imputed for between 1 and 15 percent of the former teacher items.

Table 15.   Number of source codes imputed among former teacher items, by percentage of records receiving imputation and imputation stage: 2004–05

	Number of source codes imputed, by percentage of records receiving imputation			
Imputation stage	Not imputed for any record	Imputed for 1–15 percent of the records	Imputed for 16–30 percent of the records	Imputed for more than 30 percent of the records
Stage 1	82	12	0	2
Stage 2	15	81	0	0
Stage 3	90	6	0	0

NOTE: Every question item and data entry in the questionnaires has a corresponding source code. The source codes are the 4-digit numbers found to the left of each item or data entry field in the questionnaires, which become the survey names for these data.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Former Teacher Restricted Use Data File," 2004–05.

---

[23] Source codes are used to identify specific items on the SASS and TFS questionnaires. For each questionnaire item, the four-digit source code can be found to the left of the first answer choice.

Table 16.   Number of source codes imputed among current teacher items, by percentage of records receiving imputation and imputation stage: 2004–05

	Number of source codes imputed, by percentage of records receiving imputation			
Imputation stage	Not imputed for any record	Imputed for 1–15 percent of the records	Imputed for 16–30 percent of the records	Imputed for more than 30 percent of the records
Stage 1	164	24	0	0
Stage 2	46	141	1	0
Stage 3	176	12	0	0

NOTE: Every question item and data entry in the questionnaires has a corresponding source code. The source codes are the 4-digit numbers found to the left of each item or data entry field in the questionnaires, which become the survey names for these data.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current Teacher Restricted Use Data File," 2004–05.

"Appendix H. Percentage of TFS Variables Changed During Three Stages of Imputation, by Data File" contains the total number of imputations applied at each stage to each source code.

Chapter 8. Weighting and Variance Estimation

Contained in this chapter is a discussion of the weighting and variance procedures used for the 2004–05 Teacher Follow-up Survey (TFS). The chapter discusses first the weighting procedure used to compute final weights for the interviewed teachers and next variances. Weighting is the last step in data processing. Variances are computed to estimate the reliability and are a product of the weighting procedure.

Weighting

This section describes the weighting procedure for each teacher who responded to TFS. The general purpose of weighting is to scale up the sample estimates to represent the target survey population. The steps for weighting types of respondents were similar to those used for the 2003–04 Schools and Staffing Survey (SASS). For TFS, a base weight (the inverse of the sampled teacher's probability of selection) was used as the starting point. Then, a weighting adjustment was applied that reflected the impact of the SASS teacher weighting procedure. Next, a nonresponse adjustment factor was calculated and applied using information known about the respondents from the sampling frame data. Finally, a ratio adjustment factor was calculated and applied to the sample to adjust the sample totals to frame totals in order to reduce sampling variability. The product of these factors was the final weight for TFS. See table 17 for a distribution of the final weights for the 2004–05 TFS.

Most of the steps in the weighting procedure employed weighting classes in the calculation of the weighting adjustments. Weighting classes partition the sample by key variables (such as race or age categories) and allow for differential adjustment factors to be computed for each step in the weighting procedure. This technique is especially useful when the computed factors are presumed to differ substantially, such as when patterns of nonresponse vary across subpopulations (such as by age or race). A description of how the final weight was computed as well as a brief description of each step in the weighting procedure is presented below. When computations were done within weighting classes (cells), such as nonresponse adjustments, the cells are described. Sometimes a cell did not have enough data to produce a reliable estimate; in such cases, cells were collapsed. The most important variables were always collapsed last. The collapsing criteria and the cells are described in exhibit 5.

The final TFS sample weight is the product of

> (TFS base weight) and (TFS-to-SASS weighting adjustment factor) and (TFS noninterview adjustment) and (TFS ratio adjustment)

where:

> *TFS base weight* is the inverse of the probability of selecting a teacher for TFS. This weight is the product of the SASS teacher base weight (described above) and TFS subsampling adjustment factor. The TFS subsampling adjustment factor is an adjustment that accounts for the subsampling of teachers from SASS sampled teachers. Thus, this base weight reflects the TFS probability of selection from all three stages of selection (i.e., SASS school sampling, SASS teacher sampling within school, and TFS sampling from SASS teachers).

> The *TFS-to-SASS weighting adjustment factor* is used to adjust for the fact that the SASS teacher base weights were used in selecting the TFS sample, whereas the SASS final teacher weights are

more reflective of the teacher population.[24] The weighting adjustment factor adjusts for all stages of weighting that occurred between the base and final weighting calculations for SASS teachers. For more information about the SASS teacher weighting procedure, see *Documentation for the 2003–04 Schools and Staffing Survey* (Tourkin et al. 2007).

The *TFS noninterview adjustment* is the factor used to adjust for teachers who participated in the 2003–04 SASS and were selected and determined to be eligible for the 2004–05 TFS, but did not participate in TFS. It is the weighted (product of the base weight and TFS-to-SASS weighting adjustment factor) ratio of the total eligible in-scope teachers (i.e., interviewed teachers plus noninterviewed teachers) to the total responding in-scope teachers (i.e., interviewed teachers) within cells. Variables used to define cells are presented in exhibit 5.

TFS ratio adjustment is the factor used to adjust the TFS sample totals to SASS sample totals. This adjustment ensures that the weighted number of TFS teachers (including interviews, noninterviews, and out of scopes) will be consistent with the weighted number of teachers from the 2003–04 SASS. Since the teachers who were out of scope for TFS were included in the SASS numerators, they were included in the denominators for consistency. The TFS estimates resulting from this step will not be precisely equal to SASS estimates due to the small loss of SASS teachers from eligibility for TFS due to emigration or death.

The TFS ratio adjustment is equal to the ratio of the total number of SASS teachers not selected with certainty for TFS (i.e., those teachers not automatically included in the TFS sample as mentioned in chapter 3) to the weighted TFS sample estimate of the total number of noncertainty teachers within each weighting class, or cell, defined for this step in the weighting procedure. Certainty teachers (teachers automatically included in sample for TFS based on their stratum or their measure of size) were excluded from both the numerator and denominators and were assigned a factor equal to one. Variables used to define cells are presented in exhibit 5.

Table 17.   Distribution of final weights, by data file: 2004–05

Data file	Mini-mum	Weight at given percentile									Maxi-mum	Mean
		1st	5th	10th	25th	50th	75th	90th	95th	99th		
Former teacher	2.22	6.80	11.23	15.36	33.42	64.38	127.20	230.66	345.35	1,063.77	5,819.03	125.41
Current teacher	4.25	8.84	31.92	49.29	103.86	290.62	902.19	2,095.43	2,587.98	3,598.04	7,556.59	700.90

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), "Current and Former Teacher Documentation Data Files," 2004–05.

Exhibit 5 presents the criteria for collapsing weighting classes, the definition of the weighting classes, and the collapsing ordering. Note that public charter school teachers and traditional public school teachers were adjusted separately from private school teachers. "Leavers" refer to former teachers, or teachers who left the K–12 teaching profession. "Movers" refer to teachers who moved to another school. "Stayers" refer to teachers who were still teaching at the same school as the previous year. New teachers were defined as teachers who had 3 or fewer years of teaching experience at the time of SASS (the previous year). The education (i.e., bachelor's degree or less and master's or more) and sex (male and female) categories were the same for each weighting class; but the age categories varied. The age categories were based on the distribution of the sampled teachers and designed to reduce the amount of collapsing within each weighting class. For example, the age categories used in the new public school leavers category (less

---

[24] SASS teacher weighting was not completed in time to use final teacher weights in the TFS sample selection, necessitating the use of the SASS teacher base weights in the TFS sampling.

than 24, 24–25, 26–27, 28–30, 31–34, 35–39, 40–46, 47–53, and 54 and older) were different from the age categories used in the new movers (less than 24, 24–25, 26–27, 28–30, 31–34, 35–41, and 42 and older). These age categories suggest that the new public school movers tend to be younger than the new public school leavers. The specific age categories for each weighting class are included in "Appendix I. Weighting Adjustment Cells." Cells not meeting the collapsing criteria were collapsed with other cells according to the collapsing order. Thus, cells needing collapsing were initially collapsed with cells having all other variables in common, but with an adjacent age category.

Exhibit 5.   Adjustment factors and collapsing criteria as applied to TFS weights: 2004–05

Type of teacher	Nonresponse adjustment factor		First-stage adjustment factor	
	Collapsing criteria	Collapsing order	Collapsing criteria	Collapsing order
Public school (including public charter) teachers				
New leavers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
Experienced leavers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
New movers	Factor ≤ 1.5	Education, sex, age	Factor ≥ .667 and ≤ 1.5	Age, teaching assignment, ethnicity, sex
	Interviews ≥ 15		Interviews ≥ 15	
Experienced movers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
New stayers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
Experienced stayers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
Private school teachers				
New leavers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
Experienced leavers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
New movers	Factor ≤ 1.5	Education, sex, age	Factor ≥ .667 and ≤ 1.5	Age, teaching assignment, ethnicity, sex
	Interviews ≥ 15		Interviews ≥ 15	
Experienced movers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
New stayers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	
Experienced stayers	Factor ≤ 1.5		Factor ≥ .667 and ≤ 1.5	
	Interviews ≥ 15		Interviews ≥ 15	

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), 2004–05.

Variance Estimation

In surveys with complex sample designs, such as SASS or TFS, direct estimates of sampling errors that assume a simple random sample will typically underestimate the variability in the estimates. The SASS sampling design and estimation included procedures that deviated from the assumption of simple random sampling, such as stratifying the school sample, oversampling new school teachers, and sampling with differential probabilities.

The preferred method of calculating sampling errors to reflect these aspects of the complex sample design of SASS is replication. Replication methods involve constructing a number of subsamples (i.e., replicates) from the full sample and computing the statistic of interest for each replicate. The mean square error of

the replicate estimates around the full sample estimate provides an estimate of the variance of the statistic. The replicate weights are used to compute the variance of a statistic, *Y*, as given below:

$$\text{Variance } (Y) = \left(\frac{1}{n}\right) \sum_r (Y_r - Y)^2$$

where: Y_r = the estimate of *Y* using the rth set of replicate weights; and
            n = the number of replicates.

The SASS surveys completed before 1993 used a procedure known as balanced repeated replication (BRR) for the calculation of sampling variance. BRR assumes sampling is done with replacement, and hence, BRR does not reflect the increase in precision due to sampling a high proportion of a finite population (termed the finite population correction, or FPC). For most surveys, where the sampling rates are low, the increase in precision will be small and can safely be disregarded. However, in SASS, the public sector surveys (i.e., school, principal, school district, teacher, and library media center) are designed to produce reliable state estimates. This necessarily implies high sampling rates, which can lead to very large overestimates of variance with BRR. Likewise, the private sector surveys (i.e., school, principal, and teacher) are designed to produce detailed affiliation estimates, which also imply high sampling rates and subsequent overestimation of variance with BRR.

It is possible to adjust BRR to include a finite population correction. However, since SASS uses a probability proportional to size systematic selection procedure (described in chapter 4 of *Documentation for the 2003–04 Schools and Staffing Survey* [Tourkin et al. 2007]), it is not clear what the appropriate FPC would be. It is even possible for an appropriate FPC to be greater than one, which conventional methods of variance estimation are not equipped to handle. (See Kaufman 2001.)

To overcome this limitation, a bootstrap variance estimator, which estimates the variance by simulating the sampling procedure (described in Kaufman 2001) was implemented for the 1993–94 SASS and its role was expanded in 1999–2000 and even more so in the 2003–04 SASS. The bootstrap variance estimator was used for public schools, most private schools, and public school districts in 1993–94. (See chapter 9 in *Documentation for the 2003–04 Schools and Staffing Survey* [Tourkin et al. 2007].) In 1999–2000, an additional bootstrap estimator was also included for public schools and most private schools. The bootstrap estimator used in the 2003–04 SASS was modified from the 1999–2000 estimator to make it more stable. In the 2003–04 SASS a new bootstrap estimator for both public and private school teachers was included. The bootstrap variance reflects the increase in precision due to large sampling rates because the bootstrap sampling is done systematically, without replacement, as was the original sampling.

Public schools, public school teachers, private schools sampled from the list frame, and private school teachers from schools sampled from the list frame were calculated using the updated bootstrap system. This system is based on a series of assumptions about the sampling design:

1) The traditional systematic probability proportional to size first-stage sample can be approximated using a randomized systematic sample.
2) The stratified equal probability systematic sample can be approximated by a stratified, without replacement, simple random sample.

Using these assumptions, the bootstrap replicate weights were computed from a single sample. Again, the appropriate bootstrap replicate base weights (inverse of the probability of selection) generated for the sample were subsequently reweighted by processing each set of replicate basic weights through the weighting procedure.

Since the number of certainty schools (i.e., schools guaranteed selection) was substantial, it was desirable to address the variance that results from nonresponse. Therefore, it was decided to treat nonresponse as a stage of sample selection. For certainty schools, this allowed for the reflection of a variance component that otherwise would be regarded as a bias. The nonresponse sampling model was as follows:

- For noncertainty schools, nonresponse was considered a nested random process within selected primary sampling units. That is, school nonresponse was assumed to be a random process within the random sample. Within appropriately defined cells (weighting classes), it was assumed nonresponse followed a "missing-at-random process."
- For certainty schools, nonresponse was considered the first stage of selection. It was assumed that this process followed a simple random sample without replacement model within appropriately defined cells. The frame size for this selection was assumed to be the number of selected certainty schools in the cell and the sample size was the number of responding certainty schools in the cell.

This procedure also allowed for correctly estimating variances for school-based estimates that used school teacher averages generated from the 2003–04 SASS teacher files.

To be consistent with the bootstrap procedures described above, the nonresponse modeling of certainty schools was reflected through an appropriately defined bootstrap procedure. For more details on the bootstrap methodology and how it applies to SASS, see Efron (1982), Kaufman (1992, 1993, 1994, 1998, 2001, 2002, 2003, and 2004), and Sitter (1990).

The newest version of the bootstrap procedure made it possible to compute teacher bootstrap replicate weights at the same time as the school weights, considerably reducing the processing time to form the replicates.

Each SASS data file includes a set of 88 replicate weights designed to produce variance estimates. Replicate weights were created for each of the 88 samples using the same estimation procedures used for the full sample and are included in the data files. Most of the replicate weights were produced using a bootstrap procedure. For TFS, the replicate weights were derived based on the SASS teacher replicate weights, making appropriate adjustments for the TFS sampling procedure.

As described above, the replicate weights are used to compute the variance of a statistic, Y, as given below.

$$\text{Variance } (Y) = \left(\frac{1}{88}\right) \sum_{r=1}^{88} (Y_r - Y)^2$$

where: Y_r = the estimate of Y using the rth set of replicate weights, and the number of replicate weights is 88 for SASS and TFS.

Analysis of the bootstrap replicate weights revealed that approximately 3 percent of the school (public and private) and teacher (public and private) weights fell outside a 95 percent confidence interval. These were nearly the expected 5 percent, indicating the bootstrap replicate weights were close to being normally distributed. Since the TFS replicate weights were based on the SASS teacher replicate weights, the same distribution applied.

TFS Teachers. Since the TFS sample was a proper subsample of the SASS teacher sample (i.e., TFS is representative of the whole SASS teacher sample), the SASS teacher replicates were used for the TFS sample. The TFS base weight for each TFS teacher was multiplied by each of the 88 SASS replicate

weights divided by the SASS teacher full-sample base weight for that teacher. To calculate 88 replicate weights, which should be used for variance calculations, these TFS replicate basic weights were processed through the remainder of the TFS weighting system. The replicate weights for TFS teachers are TFRPWT1 through TFRPWT88.

The SASS teacher replicate weights were generated at the same time as the school replicate weights as part of the 2003–04 bootstrap system. BRR methodology was employed rather than bootstrap if a teacher was in the private school area frame. Teacher records were assigned replicate weights by multiplying the school BRR replicate weight times the teacher's conditional probability of selection given the school was selected in the SASS school sample.

A variance estimate was obtained by first calculating the estimate for each replicate, then summing the squared deviations of the replicate estimates from the full-sample estimate, and finally dividing by the number of replicates:

$$\sum_{k=1}^{88}(\hat{y}_k - \hat{y})^2 / 88$$

where k = 1, 2 .....88;
$y_k = k^{th}$ replicate estimate; and
y = full sample estimate.

When calculating variance estimates for some small subdomains of interest (e.g., vocational education teachers), sparseness of the data may result in there being no data for some replicates. This can result in either an extremely large variance estimate or failure of the software used to calculate the variance, with possibly a warning message.

The computation of sampling errors for either TFS or SASS data using these replicate weights can be done easily with one of the following software programs: WesVar Complex Sample Software, SUDAAN (Research Triangle Institute 2001), AM Statistical Software, or STATA 9.

- *WesVar*. The user needs to create a new WesVar data file by specifying the full sample weight variable and the replicate weight variables as defined above, and the replication method, BRR. The replicate weights and the full sample weight can be highlighted and dragged to their appropriate place on the "New WesVar Data File" window. For more information, visit www.westat.com/wesvar/.
- *SUDAAN*. The user needs to specify the sample design as a "Balanced Repeated replication" design as well as the replicate weight variables. Specifying the sample design (DESIGN = BRR) is done in the procedure call statement (i.e., PROC DESCRIPT DESIGN = BRR;). The specification of the replicate weighs is done with the REPWGT statement (i.e., to produce the sampling errors for estimates from TFS data files use the statement: REPWGT TFRPWT1-TFRPWT88;). For more information, visit www.rti.org/sudaan/.
- *AM*. The user needs to set the replicate weights along with the replication method using the right-click context menu in the variable list window. Once the "Set Replicate Weights" window is displayed, the replicate weights as identified above can be highlighted and dragged into the window. At the bottom of the window are four options for replication method; BRR should be selected. For more information, visit http://am.air.org/.
- *STATA*. The use of replicate weights for the generation of standard errors is a new feature to STATA 9. First, the user needs to survey set the data (SVY SET) by defining the probability weight ([pw =]), balanced repeated replication weights (brrweight(varlist)), variance estimation

type ((vce(brr)), and turn on the mse formula (mse). Once these parameters are set, users can call up the survey settings and tell STATA which type of standard errors to produce using the SVY BRR command. SVY BRR also allows users to specify the statistics to be collected (exp_list) and the command to perform (e.g., mean or tab). For more information visit http://www.stata.com/.

This page is intentionally left blank.

Chapter 9. Reviewing the Quality of TFS Data

National Center for Education Statistics (NCES) program staff members have the responsibility of ensuring that data files are acceptable for public release. Before files are released to the public, staff members review the data for errors associated with the edit, imputation, and weighting programs. This review incorporates a number of checks including univariate, bivariate, and multivariate analyses that rigorously examine as many aspects of the data as possible without delaying timely release of Teacher Follow-up Survey (TFS) datasets.

Below are aspects of the datasets that were reviewed:

- general data quality;
- nonresponse;
- weighting; and
- external data checks, including an examination of response variance.

General Data Quality

General data quality included a number of reviews that could be characterized as consistency edits. These checks involved an examination of the individual responses, patterns of response, and summary statistics for variables and files to ensure consistency within items, respondents, and files. In addition, key variables and crosstabulations of key variables were examined for distributions and relationships that were expected based upon prior administrations and other research to check the data's face validity. The specific data checks included:

- *Edits*. The validity of the skip patterns in each TFS questionnaire was established during the processing of the data; that is, U.S. Census Bureau analysts verified that each item in each questionnaire had the number of responses it should have if skip instructions were followed correctly. Quality checks on the edit specifications were performed and resulted in some corrections (which were treated as a form of imputation).
- *Frequency counts*. Unweighted record counts for every variable were examined from the restricted-use data files. Variables with out-of-range values or inconsistent values were identified, and these values were corrected.
- *Reasonableness of data*. Univariate, bivariate, and multivariate tabulations of key survey variables were obtained and compared to estimates from the previous TFS survey. Tabulations were reviewed to determine whether the basic relationships observed were within reasonable bounds, allowing for elements of change (such as random fluctuations in variance, or a trend such as overall population growth in a state). The distributions and relationships observed were consistent with expectations.

Nonresponse

Response rates were examined for possible bias, and no evidence of bias at the unit or item level was found. The details of this analysis are discussed in greater detail in chapter 5, but the nonresponse analysis included a detailed analysis of unit nonresponse and item nonresponse.

- *Unit nonresponse*. Response rates were calculated by sector and current or former teacher status and for selected reporting characteristics. (See chapter 5 for unit response rate information.) No evidence of substantial bias was found.

- *Item nonresponse*. The extent of item nonresponse for each TFS data file was determined. (See chapter 5 for item response rate information.) Items with high nonresponse rates were identified and reported in tables. Following this review, no items were removed from the data files. However, items with a response rate lower than 70 percent are footnoted as such in published tables.

Weighting

The weighting review consisted of reviewing the distribution of TFS replicate weights. The following was done:

- For each replicate, the weights were totaled. Each replicate total, as well as the average of those numbers, was checked against the full-sample estimate. The standard errors of the replicate totals were computed and checked for reasonableness.
- A check was performed to verify that 95 percent of the replicate weights were contained in an appropriately computed 95 percent confidence interval. This was done with both the basic replicate weights and the final replicate weights.

External Data Checks

One way to verify the external validity of TFS data was to compare the total number of teachers in the 2004–05 TFS to the total number of teachers in the 2003–04 Schools and Staffing Survey (SASS). The ratio of teachers in TFS to teachers in SASS indicates whether TFS population totals were within reasonable bounds of the SASS population totals since the magnitude of the sampling error relates directly to sample size and the percentage of the universe covered by that sample.

The following three tables compare the number of TFS teachers to SASS teachers within all three school types (i.e., traditional public, public charter, and private schools) and by selected SASS teacher and school characteristics. Table 18 compares the overall number of stayer and non-stayer (i.e., movers and leavers) teachers in TFS to the total number of teachers in SASS. Table 19 compares the number of TFS teachers to the number of SASS teachers in public and private schools, and table 20 compares the number TFS teachers to the number of SASS teachers in traditional public and public charter schools.

Table 18. **Comparison of number of stayer and nonstayer teachers (both sectors) in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05**

TFS sampling strata and selected SASS teacher and school characteristics	Number of stayer teachers in TFS	Number of nonstayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
Total	3,058,779	621,427	3,680,206	3,717,998	99.0
TFS sampling strata					
TFS status					
Leaver or mover	37,343	313,689	351,031	345,838	101.5
Stayer	3,007,676	106,959	3,114,636	3,154,368	98.7
Don't know	13,760	200,779	214,539	217,792	98.5
Teaching level in SASS year					
Elementary	1,622,634	344,906	1,967,540	1,979,869	99.4
Secondary	1,436,145	276,521	1,712,666	1,738,129	98.5
Teaching experience in SASS					
Three or fewer years	512,085	161,478	673,563	618,630	108.9
More than 3 years	2,546,693	459,949	3,006,643	3,099,368	97.0
SASS teacher characteristics					
Main teaching assignment					
Early childhood/general elementary	1,094,331	207,145	1,301,476	1,305,043	99.7
Special education	341,864	93,440	435,304	432,744	100.6
Arts/music	193,261	38,146	231,407	251,385	92.1
English /language arts	306,854	63,130	369,983	385,428	96.0
ESL/bilingual education	30,857	5,741	36,598	44,480	82.3
Foreign languages	88,128	19,556	107,684	108,085	99.6
Health/physical education	178,596	31,811	210,407	197,479	106.5
Mathematics	233,592	42,759	276,351	267,592	103.3
Natural science	217,366	29,984	247,350	240,427	102.9
Social sciences	183,561	32,060	215,620	219,479	98.2
Vocational/technical education	139,005	33,508	172,513	182,290	94.6
Other	51,366	24,147	75,512	83,568	90.4
Sex					
Male	742,266	146,496	888,762	922,749	96.3
Female	2,316,513	474,931	2,791,444	2,795,249	99.9
Race/ethnicity					
White, non-Hispanic	2,623,847	516,341	3,140,188	3,113,249	100.9
Black, non-Hispanic	203,716	55,594	259,310	276,275	93.9
Hispanic	123,118	31,837	154,956	225,019	68.9
Asian, non-Hispanic	45,864	9,305	55,168	50,703	108.8
Native Hawaiian/Pacific Islander, non-Hispanic	6,479	1,036	7,515	6,913	108.7
American Indian/Alaska Native, non-Hispanic	24,862	2,699	27,561	18,916	145.7
More than one race, non-Hispanic	22,708	3,377	26,085	17,933	145.5

See notes at end of table.

Table 18.   Comparison of number of stayer and nonstayer teachers (both sectors) in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05—Continued

TFS sampling strata and selected SASS teacher and school characteristics	Number of stayer teachers in TFS	Number of nonstayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
SASS teacher characteristics—Continued					
Age					
Less than 25	147,646	42,850	190,496	155,439	122.6
25–29	365,241	126,293	491,533	473,060	103.9
30–39	730,162	141,414	871,576	900,304	96.8
40–49	838,071	122,541	960,612	959,146	100.2
50–59	863,010	144,116	1,007,126	1,062,268	94.8
60–64	87,122	35,461	122,582	130,647	93.8
65 or more	27,527	8,753	36,280	37,135	97.7
Highest degree earned					
Associate's or no degree	57,819	21,021	78,840	78,439	100.5
Bachelor's	1,570,739	328,514	1,899,253	1,910,663	99.4
Master's	1,249,945	219,322	1,469,267	1,468,290	100.1
Education specialist or certificate of advanced graduate studies	154,585	45,062	199,646	211,788	94.3
Doctorate or professional	25,691	7,508	33,199	48,817	68.0
Total school income					
Less than $35,000	871,944	238,877	1,110,821	1,064,979	104.3
$35,000–$49,999	1,282,239	238,835	1,521,073	1,546,997	98.3
$50,000–$74,999	810,588	122,497	933,085	977,858	95.4
$75,000–$99,999	85,447	19,080	104,527	118,664	88.1
$100,000 or more	8,560	2,139	10,699	9,500	112.6
SASS school characteristics					
Minority enrollment					
Less than 10 percent	981,907	161,668	1,143,576	1,071,472	106.7
10–35 percent	833,078	152,787	985,865	1,030,074	95.7
More than 35 percent	1,243,793	306,972	1,550,765	1,616,452	95.9
Census region					
Northeast	623,030	118,755	741,785	787,071	94.2
Midwest	754,749	151,040	905,789	874,744	103.5
South	1,140,325	226,054	1,366,379	1,367,920	99.9
West	540,675	125,579	666,253	688,263	96.8
Community type					
Central city	826,598	204,352	1,030,950	1,111,660	92.7
Urban fringe/large town	1,675,231	314,028	1,989,259	1,939,920	102.5
Rural/small town	556,949	103,047	659,996	666,419	99.0
School enrollment					
Fewer than 200	248,824	77,618	326,442	345,746	94.4
200–499	954,004	205,944	1,159,948	1,149,302	100.9
500–749	745,574	133,153	878,727	905,747	97.0
750–999	410,318	75,778	486,096	500,582	97.1
1,000 or more	700,058	128,934	828,992	816,622	101.5

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public Teacher and Private Teacher Data Files," 2003–04, and Teacher Follow-up Survey (TFS), "Final TFS Documentation Data Files and TFS Sample File," 2004–05.

Ninety-nine percent of the total weighted estimate of teachers in SASS is represented by the TFS sample across all school types (i.e., traditional public, public charter, and private school teachers) (table 18). For public school teachers, 98.9 percent of the SASS population estimate is captured by TFS, and for private school teachers 99.6 percent is captured (table 19).

Table 19. **Comparison of number of public and private sector school stayer and nonstayer teachers in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05**

TFS sampling strata and selected SASS teacher and school characteristics	Public sector (as identified in SASS)					Private sector (as identified in SASS)				
	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
Total	2,684,202	530,690	3,214,893	3,250,625	98.9	374,576	90,737	465,313	467,373	99.6
TFS sampling strata										
TFS status										
Leaver or mover	33,661	277,449	311,110	305,888	101.7	3,682	36,240	39,922	39,950	99.9
Stayer	2,638,061	92,426	2,730,487	2,769,522	98.6	369,615	14,534	384,149	384,846	99.8
Don't know	12,481	160,816	173,297	175,215	98.9	1,279	39,963	41,242	42,577	96.9
Teaching level in SASS year										
Elementary	1,410,247	288,549	1,698,796	1,716,404	99.0	212,387	56,357	268,744	263,465	102.0
Secondary	1,273,955	242,142	1,516,097	1,534,220	98.8	162,189	34,380	196,569	203,908	96.4
Teaching experience in SASS										
Three or fewer years	434,025	127,757	561,781	513,772	109.3	78,061	33,721	111,782	104,858	106.6
More than 3 years	2,250,178	402,934	2,653,111	2,736,852	96.9	296,516	57,016	353,531	362,516	97.5
SASS teacher characteristics										
Main teaching assignment										
Early childhood/ general elementary	953,605	174,305	1,127,911	1,131,010	99.7	140,726	32,840	173,565	174,033	99.7
Special education	325,552	87,153	412,705	414,195	99.6	16,311	6,288	22,599	18,548	121.8
Arts/music	163,661	29,622	193,283	212,606	90.9	29,600	8,524	38,124	38,779	98.3
English/language arts	268,981	54,320	323,301	339,518	95.2	37,873	8,809	46,683	45,910	101.7
ESL/bilingual education	29,772	5,609	35,381	43,190	81.9	1,085	132	1,217	1,290	94.4
Foreign languages	70,219	13,813	84,032	81,490	103.1	17,909	5,743	23,652	26,595	88.9
Health/physical education	158,510	27,486	185,996	174,808	106.4	20,086	4,324	24,411	22,671	107.7
Mathematics	201,428	36,556	237,984	231,115	103.0	32,164	6,202	38,366	36,477	105.2
Natural science	189,320	24,639	213,959	208,140	102.8	28,046	5,345	33,391	32,287	103.4
Social sciences	160,720	27,003	187,724	189,235	99.2	22,840	5,056	27,897	30,243	92.2
Vocational/technical education	133,070	31,638	164,708	175,120	94.1	5,935	1,870	7,805	7,170	108.9
Other	29,364	18,544	47,908	50,198	95.4	22,002	5,602	27,604	33,370	82.7
Sex										
Male	657,669	126,057	783,726	812,533	96.5	84,597	20,439	105,036	110,216	95.3
Female	2,026,533	404,634	2,431,167	2,438,092	99.7	289,980	70,297	360,277	357,158	100.9

See notes at end of table.

Table 19. Comparison of number of public and private sector school stayer and nonstayer teachers in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05—Continued

TFS sampling strata and selected SASS teacher and school characteristics	Public sector (as identified in SASS)					Private sector (as identified in SASS)				
	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
SASS teacher characteristics—Continued										
Race/ethnicity										
White, non-Hispanic	2,287,806	439,083	2,726,889	2,701,752	100.9	336,040	77,258	413,298	411,497	100.4
Black, non-Hispanic	192,299	50,164	242,463	257,484	94.2	11,417	5,430	16,847	18,792	89.7
Hispanic	111,052	26,744	137,795	202,463	68.1	12,067	5,094	17,160	22,556	76.1
Asian, non-Hispanic	35,062	8,031	43,092	42,167	102.2	10,802	1,274	12,076	8,536	141.5
Native Hawaiian/ Pacific Islander, non-Hispanic	5,236	933	6,169	5,969	103.3	1,243	103	1,346	943	142.7
American Indian/ Alaska Native, non-Hispanic	22,997	1,715	24,712	16,857	146.6	1,865	984	2,849	2,059	138.3
More than one race, non-Hispanic	22,203	2,991	25,194	16,322	154.4	505	386	891	1,612	55.3
Age										
Less than 25	127,357	32,880	160,237	126,549	126.6	20,289	9,969	30,259	28,889	104.7
25–29	325,044	107,917	432,961	413,766	104.6	40,196	18,376	58,572	59,293	98.8
30–39	644,979	120,940	765,919	797,714	96.0	85,183	20,474	105,657	102,590	103.0
40–49	742,315	104,672	846,988	840,342	100.8	95,756	17,869	113,625	118,804	95.6
50–59	758,511	129,139	887,650	941,650	94.3	104,499	14,977	119,476	120,618	99.1
60–64	67,312	29,614	96,926	105,364	92.0	19,809	5,846	25,656	25,283	101.5
65 or more	18,684	5,528	24,212	25,239	95.9	8,843	3,225	12,068	11,896	101.4
Highest degree earned										
Associate's or no degree	29,521	10,212	39,734	35,241	112.8	28,298	10,809	39,107	43,199	90.5
Bachelor's	1,351,664	276,254	1,627,919	1,651,425	98.6	219,074	52,260	271,335	259,238	104.7
Master's	1,141,534	196,885	1,338,419	1,330,642	100.6	108,412	22,436	130,848	137,649	95.1
Education specialist or certificate of advanced graduate studies	146,121	41,737	187,858	194,993	96.3	8,464	3,324	11,788	16,795	70.2
Doctorate or professional	15,362	5,601	20,963	38,325	54.7	10,329	1,906	12,235	10,493	116.6
Total school income										
Less than $35,000	648,240	163,553	811,793	756,374	107.3	223,704	75,323	299,028	308,606	96.9
$35,000–$49,999	1,176,840	225,654	1,402,493	1,432,828	97.9	105,399	13,181	118,580	114,169	103.9
$50,000–$74,999	768,561	120,369	888,930	937,269	94.8	42,028	2,128	44,156	40,588	108.8
$75,000–$99,999	82,229	19,020	101,250	115,018	88.0	3,218	59	3,277	3,646	89.9
$100,000 or more	8,333	2,094	10,427	9,136	114.1	228	45	273	365	74.9

See notes at end of table.

Table 19. Comparison of number of public and private sector school stayer and nonstayer teachers in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05—Continued

	Public sector (as identified in SASS)					Private sector (as identified in SASS)				
TFS sampling strata and selected SASS teacher and school characteristics	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
SASS school characteristics										
Minority enrollment										
Less than 10 percent	800,287	126,394	926,681	868,756	106.7	181,620	35,274	216,895	202,715	107.0
10–35 percent	698,645	118,720	817,365	853,818	95.7	134,433	34,067	168,500	176,256	95.6
More than 35 percent	1,185,270	285,577	1,470,847	1,528,050	96.3	58,523	21,395	79,918	88,402	90.4
Census region										
Northeast	520,291	97,565	617,856	662,348	93.3	102,739	21,190	123,929	124,723	99.4
Midwest	666,738	130,042	796,779	768,491	103.7	88,012	20,998	109,010	106,253	102.6
South	1,017,142	192,719	1,209,861	1,214,670	99.6	123,183	33,335	156,518	153,251	102.1
West	480,031	110,365	590,396	605,117	97.6	60,643	15,214	75,857	83,146	91.2
Community type										
Central city	680,843	172,502	853,345	929,391	91.8	145,756	31,850	177,605	182,269	97.4
Urban fringe/large town	1,481,358	266,197	1,747,555	1,704,231	102.5	193,873	47,831	241,704	235,689	102.6
Rural/small town	522,002	91,991	613,993	617,003	99.5	34,948	11,056	46,003	49,416	93.1
School enrollment										
Fewer than 200	117,491	28,791	146,282	167,473	87.3	131,333	48,827	180,160	178,273	101.1
200–499	814,215	175,904	990,118	978,881	101.1	139,789	30,040	169,830	170,420	99.7
500–749	703,942	126,527	830,469	850,272	97.7	41,632	6,626	48,258	55,475	87.0
750–999	373,228	72,741	445,969	466,332	95.6	37,091	3,036	40,127	34,250	117.2
1,000 or more	675,327	126,727	802,054	787,667	101.8	24,731	2,207	26,938	28,955	93.0

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public Teacher and Private Teacher Data Files," 2003–04, and Teacher Follow-up Survey (TFS), "Final TFS Documentation Data Files and TFS Sample File," 2004–05.

Comparisons were made by TFS sampling strata as well as by selected SASS teacher and school characteristics. These identified a few areas where the TFS population differed from the SASS population (tables 19 and 20). The TFS sample of public school teachers (including public charter) contained a larger proportion of non-Hispanic American Indian/Alaska Native and non-Hispanic multirace racial/ethnic groups and a smaller proportion of teachers who were Hispanic and teachers who had doctorate or professional degrees (table 19). Comparisons of private school teachers revealed a larger proportion of non-Hispanic Asian, non-Hispanic Native Hawaiian/Pacific Islander, and non-Hispanic American Indian/Alaska Native racial/ethnic groups and a smaller proportion of the non-Hispanic multirace racial/ethnic group (table 19). Among only public charter school teachers who participated in TFS, there was a larger proportion of the non-Hispanic American Indian/Alaska Native racial/ethnic group and teachers whose total school income was in the range of $75,000 to $99,999. There was a smaller proportion of public charter school teachers who were 60 to 64 years old, had doctorate or professional degrees, and whose total school income was in the range of $50,000 to $74,999 (table 20).

These imbalances are not unusual since TFS only controls for status (stayers, movers, leavers, and unknown), within school type (traditional public, public charter, and private), experience groups (new/experienced), teacher's grade level (elementary/middle/secondary), and minority status (minority/nonminority). The uncontrolled variables tend to have high variances and produce random fluctuations.

Table 20.   Comparison of number of traditional public and public charter school stayer and nonstayer teachers in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05

TFS sampling strata and selected SASS teacher and school characteristics	Traditional public (as identified in SASS)					Public charter (as identified in SASS)				
	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
Total	2,654,960	518,461	3,173,421	3,208,486	98.9	29,242	12,229	41,472	42,139	98.4
TFS sampling strata										
TFS status										
Leaver or mover	33,450	273,020	306,470	301,374	101.7	211	4,428	4,639	4,514	102.8
Stayer	2,609,380	90,794	2,700,174	2,738,194	98.6	28,681	1,631	30,312	31,328	96.8
Don't know	12,130	154,646	166,776	168,917	98.7	351	6,170	6,520	6,297	103.5
Teaching level in SASS year										
Elementary	1,390,913	282,023	1,672,936	1,691,179	98.9	19,334	6,526	25,860	25,225	102.5
Secondary	1,264,047	236,438	1,500,485	1,517,307	98.9	9,908	5,704	15,612	16,913	92.3
Teaching experience in SASS										
Three or fewer years	424,182	122,475	546,657	498,284	109.7	9,843	5,281	15,124	15,488	97.6
More than 3 years	2,230,778	395,986	2,626,764	2,710,202	96.9	19,400	6,948	26,348	26,650	98.9
SASS teacher characteristics										
Main teaching assignment										
Early childhood/general elementary	940,582	169,674	1,110,256	1,113,193	99.7	13,024	4,631	17,655	17,817	99.1
Special education	322,569	86,236	408,804	410,862	99.5	2,984	917	3,901	3,334	117.0
Arts/music	162,007	28,428	190,435	209,967	90.7	1,654	1,195	2,849	2,638	108.0
English/language arts	266,086	53,408	319,494	335,588	95.2	2,894	912	3,807	3,930	96.9
ESL/bilingual education	29,515	5,577	35,092	42,852	81.9	257	32	289	338	85.6
Foreign languages	69,212	13,065	82,277	79,995	102.9	1,007	748	1,754	1,495	117.3
Health/physical educ.	158,105	27,214	185,319	173,031	107.1	405	272	677	1,777	38.1
Mathematics	199,288	35,673	234,961	227,810	103.1	2,140	883	3,023	3,304	91.5
Natural science	187,765	23,908	211,673	205,781	102.9	1,555	732	2,286	2,359	96.9
Social sciences	158,198	26,119	184,317	186,632	98.8	2,522	884	3,407	2,604	130.9
Vocational/technical education	133,070	31,081	164,151	174,006	94.3	0	558	558	1,114	50.1
Other	28,563	18,079	46,643	48,769	95.6	801	465	1,266	1,429	88.6
Sex										
Male	650,605	121,974	772,579	801,148	96.4	7,064	4,083	11,147	11,385	97.9
Female	2,004,355	396,487	2,400,841	2,407,338	99.7	22,178	8,147	30,325	30,753	98.6

See notes at end of table.

Table 20.   Comparison of number of traditional public and public charter school stayer and nonstayer teachers in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05—Continued

TFS sampling strata and selected school and teacher characteristics	Traditional public (as identified in SASS)					Public charter (as identified in SASS)				
	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
SASS teacher characteristics— Continued										
Race/ethnicity										
White, non-Hispanic	2,266,555	430,780	2,697,335	2,672,192	100.9	21,251	8,303	29,554	29,560	100.0
Black, non-Hispanic	187,963	48,427	236,389	251,079	94.1	4,336	1,737	6,074	6,404	94.8
Hispanic	108,428	25,432	133,860	198,188	67.5	2,623	1,312	3,935	4,275	92.0
Asian, non-Hispanic	34,347	7,887	42,234	41,364	102.1	715	144	859	802	107.0
Native Hawaiian/Pacific Islander, non-Hispanic	5,236	823	6,060	5,820	104.1	0	109	109	150	72.8
American Indian/Alaska Native, non-Hispanic	22,730	1,474	24,204	16,489	146.8	267	241	508	368	137.9
More than one race, non-Hispanic	22,203	2,649	24,851	15,969	155.6	0	342	342	352	97.1
Age										
Less than 25	125,839	31,528	157,367	123,718	127.2	1,518	1,353	2,870	2,831	101.4
25–29	319,347	105,221	424,569	404,367	105.0	5,697	2,696	8,392	9,399	89.3
30–39	634,374	116,774	751,148	784,087	95.8	10,605	4,166	14,772	13,627	108.4
40–49	736,051	103,084	839,135	832,467	100.8	6,265	1,588	7,853	7,875	99.7
50–59	753,672	127,096	880,769	934,422	94.3	4,839	2,043	6,881	7,227	95.2
60–64	67,312	29,316	96,628	104,585	92.4	0	298	298	779	38.2
65 or more	18,364	5,442	23,806	24,840	95.8	319	86	405	399	101.5
Highest degree earned										
Associate's or no degree	28,853	9,659	38,512	33,878	113.7	668	553	1,221	1,362	89.6
Bachelor's	1,333,170	269,000	1,602,170	1,624,390	98.6	18,494	7,254	25,748	27,035	95.2
Master's	1,132,753	193,383	1,326,135	1,319,103	100.5	8,781	3,503	12,284	11,539	106.5
Education specialist or certificate of advanced graduate studies	144,878	41,063	185,941	193,236	96.2	1,242	675	1,917	1,757	109.1
Doctorate or professional	15,305	5,356	20,662	37,879	54.5	57	245	302	446	67.7
Total school income										
Less than $35,000	636,876	156,288	793,165	736,836	107.6	11,364	7,265	18,629	19,538	95.3
$35,000–$49,999	1,162,417	221,556	1,383,973	1,415,783	97.8	14,423	4,097	18,520	17,044	108.7
$50,000–$74,999	765,881	119,502	885,383	931,934	95.0	2,680	867	3,547	5,335	66.5
$75,000–$99,999	81,453	19,020	100,473	114,797	87.5	776	0	776	221	350.9
$100,000 or more	8,333	2,094	10,427	9,136	114.1	0	0	0	0	0.0

See notes at end of table.

Table 20.   **Comparison of number of traditional public and public charter school stayer and nonstayer teachers in 2004–05 TFS population to teachers in 2003–04 SASS, by TFS sampling strata and selected SASS teacher and school characteristics: 2003–04 and 2004–05—Continued**

TFS sampling strata and selected school and teacher characteristics	Traditional public (as identified in SASS)					Public charter (as identified in SASS)				
	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS	Number of stayer teachers in TFS	Number of non-stayer teachers in TFS	Total number of teachers in TFS	Total number of teachers in SASS	Ratio of total number of teachers in TFS to total in SASS
SASS school characteristics										
Minority enrollment										
Less than 10 percent	796,938	125,183	922,121	864,360	106.7	3,350	1,210	4,560	4,396	103.7
10–35 percent	691,510	114,872	806,382	842,362	95.7	7,135	3,848	10,983	11,456	95.9
More than 35 percent	1,166,512	278,405	1,444,918	1,501,764	96.2	18,758	7,171	25,929	26,286	98.6
Census region										
Northeast	512,717	95,686	608,403	654,940	92.9	7,574	1,879	9,454	7,408	127.6
Midwest	659,220	127,089	786,309	758,386	103.7	7,517	2,953	10,470	10,105	103.6
South	1,012,086	189,256	1,201,343	1,204,104	99.8	5,056	3,462	8,518	10,566	80.6
West	470,937	106,430	577,366	591,057	97.7	9,095	3,935	13,030	14,060	92.7
Community type										
Central city	666,000	165,665	831,665	906,496	91.7	14,843	6,837	21,680	22,895	94.7
Urban fringe/large town	1,470,339	261,959	1,732,299	1,688,535	102.6	11,019	4,237	15,256	15,696	97.2
Rural/small town	518,621	90,836	609,457	613,456	99.3	3,381	1,155	4,536	3,547	127.9
School enrollment										
Fewer than 200	112,333	25,054	137,386	157,615	87.2	5,158	3,738	8,896	9,858	90.2
200–499	804,073	170,407	974,480	963,866	101.1	10,142	5,497	15,638	15,015	104.1
500–749	697,807	125,500	823,307	842,581	97.7	6,135	1,027	7,163	7,690	93.1
750–999	369,552	71,691	441,243	462,013	95.5	3,675	1,050	4,726	4,319	109.4
1,000 or more	671,195	125,810	797,004	782,411	101.9	4,132	918	5,050	5,256	96.1

NOTE: Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), "Public Teacher and Private Teacher Data Files," 2003–04, and Teacher Follow-up Survey (TFS), "Final TFS Documentation Data Files and TFS Sample File," 2004–05.

Chapter 10. Information on Data Files and Merging Components

The 2004–05 Teacher Follow-up Survey (TFS) is composed of two survey questionnaires—the Questionnaire for Current Teachers and the Questionnaire for Former Teachers. The questionnaires were administered to a sample of teachers from public (including public charter) and private schools that responded to the 2003–04 Schools and Staffing Survey (SASS). These two questionnaires became two data files that followed the populations targeted by the questionnaires: the current teacher data file, which includes teachers who remained in the same school as during the SASS school year (stayers) and teachers who moved to a new school in the 2004–05 school year (movers), and the former teacher data file, which includes teachers who left the pre-K–12 teaching profession after the 2003–04 school year (leavers). Table 21 identifies each data file and the questionnaire data used to build the file.

Table 21.   Names of data files and the questionnaires from which the data were drawn: 2004–05

Data file	Questionnaire source
Current teacher	Questionnaire for Current Teachers
Former teacher	Questionnaire for Former Teachers

SOURCE: U.S. Department of Education, National Center for Education Statistics, Teacher Follow-up Survey (TFS), 2004–05.

Availability of Data

The 2004–05 TFS raw data are available in a restricted-use form only. The restricted-use files include confidentiality edits, which add "noise" to the data in order to make the identification of respondents in published data less certain. (See the section below on "Confidentiality Edits to the Data.") Access to restricted-use data files is limited to individuals associated with organizations that have received a license to use SASS and TFS data. How to receive a restricted-use license is discussed in the next section.

Restricted-use data are accessed through an ECB, which is a searchable codebook, or data dictionary, on a CD-ROM that produces data files as specified by the user. Data are restricted-use because they contain identifiable information, which is confidential and protected by law. While direct identifiers, such as the respondent's name, are not included on the files, the restricted-use files do feature more variables that can indirectly identify a respondent or that can be used to link SASS and TFS with the Common Core of Data (CCD) or other data files, which could provide the name of the school and lead to the identification of individual respondents.

The 2004–05 TFS restricted-use data are released in accordance with the provisions of the Education Sciences Reform Act of 2002, Public Law 107-279 (20 U.S.C. 9873), the Privacy Act of 1974, the USA Patriot Act of 2001, and the E-Government Act of 2002. NCES is responsible for protecting the confidentiality of individual respondents and releases data (CD-ROMs) for statistical purposes only. Record matching or deductive disclosure by any user is prohibited by federal law.

How to Get Restricted-Use Data Files

Researchers who can demonstrate a need for more detailed information may request access to the restricted-use datasets for statistical research purposes, provided that they follow computer security requirements and fill out an Affidavit of Nondisclosure.

Researchers requesting access to the restricted-use datasets must obtain a license to use those data by providing the following information:

- the title of the survey(s) to which access is desired (e.g., 2004–05 Teacher Follow-up Survey to the 2003–04 Schools and Staffing Survey);
- a detailed discussion of the statistical research project that necessitates accessing the NCES survey;
- the name of the principal project officer at the institution who will be heading up the research effort and who will enforce the legal provisions of the license agreement;
- the number, name(s), and job title(s) of professional and technical staff, including graduate students, who will be accessing the survey dataset; and
- the estimated loan period necessary for accessing the NCES survey dataset.

Return all of the above information to

>NCES Data Security Office
>Department of Education/NCES/ODC/SSP
>1990 K Street NW
>Room 9061
>Washington, DC 20006

All of these procedures are detailed in the *NCES Restricted-Use Data Procedures Manual*, available for download at http://nces.ed.gov/statprog/confid6.asp.

After the access request has been reviewed, the requestor will be informed whether a license to use the restricted data has been approved.

Requestors and/or institutions that violate the agreement are subject to a fine of not more than $250,000 (under the provisions of 18 U.S.C. 3559 and 3571) or imprisonment for not more than 5 years, or both. The confidentiality provisions that NCES must follow by law can be found at http://nces.ed.gov/statprog.

Researchers requesting TFS data will also receive the SASS data files from the previous year.

Understanding the Data Files

Confidentiality Edits to the Data

The restricted-use data files have been altered according to NCES standards. Known as confidentiality edits, "noise" was added to the data in order to make the identification of respondents less certain. These edits directly alter some data for individual respondents, but preserve the overall distributions and level of detail in all variables included on the file. There are several ways in which data can be altered, including blanking and imputing for randomly selected records; blurring (e.g., combining multiple records through some averaging process into a single record); adding random noise; and data swapping or switching (e.g., switching the variable for age from a predetermined pair of individuals). Both restricted-use TFS data files (i.e., Current Teacher and Former Teacher) were altered through one or more of these methods.

Treatment of Public Charter Schools and Schools Funded by the Bureau of Indian Affairs

Teachers who taught in schools funded by the Bureau of Indian Affairs (BIA) and public charter schools in the 2003–04 school year were included in the SASS sample. While teachers from public charter schools were included in TFS, those from BIA-funded schools were excluded from the data collection. BIA-funded school teachers were not included because of insufficient sample sizes in TFS. In SASS, the

data from BIA-funded schools, principals, teachers, and school library media centers were placed on separate data files that only include BIA-funded school-related components.

Public charter schools were first included in the 1999–2000 administration of SASS. At that time, the number of public charter schools was small enough that all known to be operational in 1998–99 and still operating in 1999–2000 were surveyed. The number of public charter schools has continued to grow, making it more feasible to sample public charter schools. A sample of 303 public charter schools was selected for the 2003–04 SASS. (See chapter 4 in the *Documentation for the 2003–04 Schools and Staffing Survey* [Tourkin et al. 2007] for details.) Data from these respondents were included in the public sector SASS and TFS data files. The variable CHARFLAG, which identifies whether the public school is a traditional public school or a public charter school, can be used for separately analyzing public charter data.

There were instances when schools did not fit exclusively into a single category of traditional public, public charter, or BIA-funded schools. In these instances, the following criterion for determining school type was applied:

- Schools included on the BIA Directory of schools were categorized as BIA-funded schools and included on the BIA SASS files.
- Schools that were on the BIA Directory of schools but also indicated that they were public charter schools were categorized as BIA-funded schools and included on the BIA SASS files.

In addition, how a school was classified on CCD (as public, public charter, or BIA) may not match how the school classified itself on the questionnaire. The following decisions were made to assign the school's type in SASS:

- Schools that were classified as public charter schools on CCD but did not claim charter school status on the questionnaire were categorized as traditional public schools.
- Schools that were classified as public schools on CCD but claimed to be charter schools on the questionnaire were categorized as public charter schools.
- Schools were not asked on the questionnaire whether they were funded by BIA; there were no inconsistencies with the school's type as it was assigned on the SASS sampling frame or on the SASS data files.

Categories of Variables

Variables on TFS data files were organized into the following five categories on each record layout: frame, survey, created, weighting, and imputation flag variables. Each of these categories was further separated into subcategories that provide more detail on each variable's source. The purpose of these categories is to help the user better understand what types of variables are included on the files and what the sources were for the variables.

Variables were classified as frame variables if they were drawn from or based on the TFS sampling frame, the SASS sampling frame, the CCD, or the Private School Universe Survey (PSS). There are four types of frame variables, or subcategories, identifying the source of each frame variable: TFS frame, 2003–04 SASS frame, 2001–02 CCD or PSS, or 2003–04 CCD or PSS. Frame variables used in the SASS or TFS sampling operations are explained in greater detail in chapter 3. Variables that were not used for sampling purposes but were classified as frame variables and placed on the files were selected because they provided potentially valuable information to the user that was not available from the survey itself. Examples of frame variables include the respondent's control, or identification, number (i.e.,

CNTLNUMS for schools and CNTLNUMT for teachers) and locale codes (i.e., SLOCP_03 and URBANS03). The frame variables are listed in "Appendix J. Frame and Created Variables."

Survey variables are the actual variables drawn from the questionnaire responses. Each item on a questionnaire has a small number printed to the left. This series of numbers is the source code. A single letter was added to the beginning of the series to signify which type of respondent (i.e., teacher, school) is associated with a source code. All source codes for TFS begin with the letter "F." For example, on the Questionnaire for Current Teachers, item 1a has the source code 0050 printed to the left. On the data file, the source code for this item is F0050. In SASS, the letter "D" was added for district, "S" for school, "A" for principal or administrator, "T" for teacher, and "M" for school library media center.

Created variables are based on survey variables, frame variables, other created variables, or a combination of these. These variables are frequently used in NCES publications and have been added to the files to facilitate data analysis. The code used to create these variables can be found in the description of each variable in the Codebook Window of the restricted-use Electronic Codebook. There are two subcategories for created variables based on whether the data used to create the variable were from "within" the teacher's record (either from SASS or TFS data files) or from "other" data files in SASS. For example, the variable TOTEXPER on the current and former teacher data files is categorized as a "within" created variable, because it identifies the teacher's total years of teaching experience as reported on the SASS teacher questionnaire. Likewise, STTUS_TF is included on the current and former teacher data files and is considered a "within" created variable, because it identifies whether the respondent has continued to teach at the same school as the previous year (stayer), has moved to a new school for the 2004–05 school year (mover), or has stopped teaching in grades pre-K–12 (leaver), as reported on the TFS surveys. A created variable labeled as being "other" is based on data from a file other than the teacher files. "Other" created variables were typically based on the SASS school or district data files and then placed on the TFS data files as a convenience to the user. For example, ENRK12UG is found on both the current and former teacher data files and is classified as an "other" created variable, because it provides the total student enrollment of the SASS school, as reported on the SASS school survey. The created variables are listed in "Appendix J. Frame and Created Variables."

There are two types of weighting variables on each file. (For more information on weighting and standard errors, see chapter 8.) The first is the sampling weight, or final weight for the respondent, and the second includes the 88 replicate weights. The final weight adjusts for nonresponse and oversampling and is used so that estimates represent the population rather than simply the sample. The replicate weights are used as a set to generate standard errors for estimates. On the TFS files, the final weight is called TFSFINWT and the replicate weights are TFRPWT1 through TFRPWT88.

The imputation flags identify whether a survey item was imputed for missing data (described in detail in chapter 7) or whether a created variable was imputed because of a nonresponding school or district. All survey variables have a corresponding imputation flag that indicates whether a value was imputed and, if so, what method was used. All survey imputation flags begin with "F_" and are followed by the name of the variable. For example, the imputation flag for F0601 from the teacher files is F_F0601. In addition, there are three frame variables (SLOCP_03, STATE, and SC_ZIP) that have a corresponding imputation flag (FL_SLC03, FL_STAT, and FL_ZIP). The variable SLOCP_03 only applies to TFS respondents who taught in a public sector school in the 2003–04 school year. This variable and its flag were pulled directly from the 2003–04 CCD. The imputation flags for STATE and SC_ZIP identify the source of the data when they were imputed, so that the user is aware of whether the school's or the respondent's address was used to identify the respondent's current state and ZIP code.

Certain created variables on TFS were also given imputation flags. The imputation flag for these created variables indicates whether the school or district failed to respond to SASS and, if so, then what type of

imputation was used as the source for the data. If the school or district did not respond to SASS, data are still present for these variables on the TFS files. All created variable imputation flags begin with "FL_" and are followed by at least the beginning of the name of the created variable. For example, the variable ENRK12UG comes from the SASS school file and provides the total K–12 and ungraded enrollment in the base-year school. It is placed on both the current and former teacher data files. The variable's imputation flag is called FL_ENRK. These created variables were built with data from either the SASS district or school data files and placed on the TFS current and former teacher data files. However, if the district or school failed to respond to SASS, data would not be available to place on other files. These data were imputed using data from the sampling frame, if available, or imputed by hand.

Nonresponding Units

As described in chapter 3 on sample selection, the school was the primary sampling unit for SASS. For each sampled school, the principal, selected teachers, and the library media center and district, if applicable, were included in SASS. Not all of these types of respondents chose to participate in SASS. Consequently, it is possible to have several teacher records but no corresponding school record, because the school did not complete the SASS school questionnaire. Similarly, the district could have agreed to allow its schools to participate in SASS but failed to complete its questionnaire, resulting in having completed questionnaires for schools and principals but no corresponding district data. Table 22 below identifies the number of cases that have a corresponding unit that did not respond to the 2003–04 SASS. This information is particularly useful when match merging data files to identify how many cases are missing.

Table 22.   Number of missing cases in combined SASS datasets, by nonresponding component and dataset providing unit of analysis: 2003–04

Unit of analysis	Observations	Nonresponding public component			
		Public school districts	Principals	Schools	School library media centers
Public school principal	8,143	1,288	†	407	1,249
Public school	7,991	1,221	255	†	1,213
Public school teacher	43,244	6,637	2,166	2,965	5,607
Public school library media center	7,229	1,126	335	451	†

Unit of analysis	Observations	Nonresponding private component			
		†[1]	Principals	Schools	†[1]
Private school principal	2,376	†	†	88	†
Private school	2,456	†	168	†	†
Private school teacher	7,979	†	509	475	†

Unit of analysis	Observations	Nonresponding BIA-funded component[2]			
		†[3]	Principals	Schools	School library media centers
BIA-funded school principal	146	†	†	5	24
BIA-funded school	145	†	4	†	23
BIA-funded school teacher	624	†	21	30	81
BIA-funded school library media center	124	†	2	2	†

† Not applicable.
[1] Private sector components did not include the School District Questionnaire and private schools did not receive the School Library Media Center Questionnaire.
[2] BIA refers to the Bureau of Indian Affairs.
[3] BIA-funded sector components did not include the School District Questionnaire.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), 2003–04.

Table 23 provides similar information for TFS. The TFS data files are divided by sector (i.e., current public school teacher, current private school teacher, former public school teacher, and former private school teacher data files) and match merged with each SASS data file to identify which TFS cases do not have a matching component for analysis. Districts, principals, schools, and library media centers that did not have a teacher who was sampled for and responded to TFS are excluded from this table.

Table 23.   Number of missing cases in combined datasets, by merged SASS data file and TFS data file by sector providing unit of analysis: 2003–04 and 2004–05

		Nonresponding SASS public component			
Unit of analysis	Observations	Public school districts	Principals	Schools	School library media centers
Public school current teacher	3,497	283	143	187	396
Public school former teacher	1,826	181	94	115	243
		Nonresponding SASS private component			
Unit of analysis	Observations	†[1]	Principals	Schools	†[1]
Private school current teacher	1,279	†	54	53	†
Private school former teacher	827	†	59	50	†

† Not applicable.
[1] Private schools did not receive the School District Questionnaire or the School Library Media Center Questionnaire.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Schools and Staffing Survey (SASS), 2003–04; Teacher Follow-up Survey (TFS), 2004–05.

Linking Files Within and Across TFS and SASS

In SASS, when each school was selected for the school sample, its principal or school head was selected for the principal sample, along with a sample of teachers at that school. For public schools, the school district having jurisdiction over the sampled school was selected for the district sample. For public and BIA-funded schools, a staff member who was responsible for the school library media center, if any, was also included in the school library media center sample. The School Library Media Center Questionnaire was not given to private schools. In the following school year, a sample of teachers from those who responded to the teacher questionnaire in SASS were contacted to complete TFS.

On restricted-use TFS and SASS data files, any combination of the school, principal, teacher, and library media center (if applicable) datasets within each school sector can be merged using the school's control number (CNTLNUMS). Consequently, all of the SASS files can be merged with the TFS data files. The one exception is that no teachers who taught in a BIA-funded school during the SASS school year were included in the 2004–05 TFS. The school control number is present on all TFS and SASS files, except for the district data file,[25] and will link them together. The SASS and TFS teacher data files can be linked to provide data on two school years using the teacher's control number (CNTLNUMT).

The public teacher, school, principal, and library media center datasets may be merged with the district dataset. School and district datasets can be merged using the district's control number (CNTLNUMD) or by parsing out the first five digits of the school's control number and the district's control number.

There are two ways in which files can be merged. The first involves match merging files. An example of this is when the user would like to merge a school's record with the records of its teachers. The school

---

[25] The school control number is not included on the district data file because usually there is more than one school associated with a district. Consequently, the school's control number cannot be used to merge the district and school data files.

and the teachers are linked through the school's control number. Instructions on how to match merge files in SAS, SPSS, and Stata are provided below. The second method of merging is appending, or concatenating, files. For example, if a user would like to analyze the current and former teacher data, these two files can be appended together. Because these files do not need to be "matched," no control number needs to be specified to append the files. This type of merging is not discussed in this chapter. Refer to the manual for the statistical program being used to determine how to append files and for additional information on how to merge files.

Sample SAS Syntax for Merging Data Files and Attaching Variable Labels

Merging Restricted-Use Data Files Using the Teacher Control Number (CNTLNUMT)

When merging records for teachers from the TFS data files to their records in the SASS data files, the teacher's control number, CNTLNUMT, is used to match files. In the SAS syntax below, please note that both data files being merged must be sorted by the variable listed in the "by" statement prior to performing the merge. Comments to explain lines of syntax are contained within "/* */"; words in italics are meant to be replaced by meaningful file or variable names.

```
proc sort data = TFSdataset;
by CNTLNUMT;
run;
proc sort data = SASSdataset;
by CNTLNUMT;
run;
data newfilename;                        /* create new merged file name */

merge TFSdataset (in=a) SASSdataset;     /* merge the two files and specify the TFS dataset as
                                            unit of analysis */
by CNTLNUMT;
if a=1;                                  /* keep all TFS records and only the matching SASS
                                            records */

run;
```

Merging Restricted-use Data Files Using the School Control Number (CNTLNUMS)

When merging any of the school, principal, teacher, or school library media center data files together for a given school, the school's control number, CNTLNUMS, is used to match data files. In the SAS syntax below, please note that both data files being merged must be sorted by the variable listed in the "by" statement prior to performing the merge. Comments to explain lines of syntax are contained within "/* */"; words in italics are meant to be replaced by meaningful file or variable names.

```
proc sort data = dataset1;
by CNTLNUMS;
run;
proc sort data = dataset2;
by CNTLNUMS;
run;
data newfilename;                /* create new merged file name */
merge dataset1 (in=a) dataset2;  /* merge the two files and specify dataset1 as unit of
                                    analysis */

by CNTLNUMS;
```

if a=1;   /* keep all dataset1 records and only matching dataset2
                                                                records */

run;

Merging the Restricted-use Public School District Data File with Other Public Sector Data Files

There are two ways to merge the public school district data file with other public sector data files. The first is with the district's control number (CNTLNUMD). This variable is included on the public school district data file as well as the public school data file. The sample syntax provided above is correct, except that the merging variable will be CNTLNUMD.

The second method is by parsing out the first five digits of the district's and the school's control number. Users will need to use this method if the school did not respond to SASS. The first five digits of CNTLNUMS and CNTLNUMD are identical, so users can create a new variable using a substring of these control numbers and merge the data files by the new variable name. The SAS syntax provided below illustrates how to merge the public school district data file with other data files using a substring. Please note that the data files being merged must be sorted by the variable listed in the "by" statement prior to performing the merge. Comments to explain lines of syntax are contained within "/* */"; words in italics are meant to be replaced by meaningful file or variable names.

data *workfilename1*;
set *school_or_principal_or_teacher_pr school library media centerdatafile*;
newvariablename = **substr (**CNTLNUMS,1,5**);**
run;
proc sort;
by *newvariablename*;
run;
data *workfilename2*;
set *districtdatafile*;
newvariablename = **substr (**CNTLNUMD,1,5**);**
run;
proc sort;
by *newvariablename*;
run;
data *mergedfilename*;
merge *workfilename1 workfilename2*;   /* No unit of analysis file is identified, so all records
                                                                   from both files will remain */
by *newvariablename*;
run;

Attaching Value Labels to Variables Extracted from the Electronic Codebook (ECB)

While the formatting syntax is provided, it is up to the user to call up the labels. There are three primary ways to accomplish this. First, value labels for each variable can be reviewed within the ECB. When variables are extracted from the ECB there is a box on the right-hand side of the pop-up window that requests the creation of a codebook. Check this box in order to have the ECB create a text file with the codebook information for all extracted variables. Then within this text file use the find function to locate the variable and determine the value labels.

Second, labels may be manually attached using the PROC step in SAS. To do this, review the syntax created from the extraction process to determine the value label name associated with each variable. In

general, the Format name drops the last digit or letter in the variable name and adds the letter "F" at the end. There are exceptions to this rule.

As an example, the appropriate SAS syntax for a PROC FREQ is given below. Words in italics are meant to be replaced by meaningful file or variable names.

> **proc freq;**
> **format** *varname valuename***.;**
> **tables** *varname*;
> **run;**

A third method is to create a permanent value label library in SAS. This requires users to manipulate the SAS syntax generated from the extraction. To begin, users need to create a permanent library for the value formats that includes all of the value formats they would like to keep. In the SAS syntax that follows, please note that comments to explain lines of syntax are contained within "/* */" and words in italics are meant to be replaced by meaningful file or variable names.

> libname library 'C:*librarypath*';   /*assigns format library, libname must be "library"*/
>
> **proc format** library=library;   /*creates permanent formats in the directory specified
>   in library libname statement*/
> [List all of the value formats here]
> VALUE   URBANIF
> **1** = "Large or mid-size central city"
> **2** = "Urban fringe of large or mid-size city"
> **3** = "Small town/Rural"
> ;
> VALUE   VIOLPRF
> **0** = "School does not have a violence prevention program"
> **1** = "School has a violence prevention program but no formal procedure for assessing its
>   effectiveness"
> **2** = "School has a violence prevention program and a formal procedure for assessing its
>   effectiveness";

The above syntax is written before the user's first data step and set statements. Within the data step programming that follows, the following format commands must be included:

> FORMAT *varname valuename*.;

Sample SPSS Syntax for Merging Data Files

NOTE: Both data files being merged must be sorted by the variable listed in the "by" statement prior to performing the merge. In SPSS, value labels are attached automatically during the extraction process.

Merging Restricted-use Data Files Using the Teacher Control Number (CNTLNUMT)

When merging records for teachers from the TFS data files to their records in the SASS data files, the teacher's control number, CNTLNUMT, is used to match files. The SPSS syntax is provided below. Words in italics are meant to be replaced by meaningful file or variable names.

```
get file = 'dataset1.sav';
sort cases by CNTLNUMT(A);
save outfile = 'dataset1.sav';
get file = 'dataset2.sav';
sort cases by CNTLNUMT(A);
save outfile = 'dataset2.sav';
match files file = 'dataset1.sav';
        /table 'dataset2'
        /by CNTLNUMT.
save outfile = 'mergeddatafile.sav';
```

Merging Restricted-use Data Files Using the School Control Number (CNTLNUMS)

When merging any of the school, principal, teacher, or library media center files together for a given school, the school's control number, CNTLNUMS, is used to match files. The SPSS syntax is provided below. Words in italics are meant to be replaced by meaningful file or variable names.

```
get file = 'dataset1'.
sort cases by CNTLNUMS(A).
save outfile = 'dataset1'.
get file = 'dataset2'.
sort cases by CNTLNUMS(A).
save outfile = 'dataset2'.
match files file = 'dataset1'
        /table 'dataset2'
        /by CNTLNUMS.
save outfile = 'mergeddatafilelocation'.
```

Merging Restricted-use Public School District File with Other Public Sector Files

There are two ways to merge the district file with other public sector files. The first is with the district's control number (CNTLNUMD). This variable is included on the district file as well as the public school file. The sample code provided above is correct, except that the merging variable will be CNTLNUMD.

The second method is by parsing out the first five digits of the district's and the school's control number. Users will need to use this method if the school did not respond to SASS. The first five digits of CNTLNUMS and CNTLNUMD are identical, so users can create a new variable using a substring of these control numbers and merge the files by the new variable name. The SPSS syntax provided below illustrates how to merge the public school district file with other data files using a substring. Please note that the data files being merged must be sorted by the variable listed in the "by" statement prior to performing the merge. In addition, including "(a5)" for the substring specifies the new variable as a five-character alphanumeric variable. Words in italics are meant to be replaced by meaningful file or variable names.

```
get file = 'school_or_principal_or_teacher_or_librarydatafile.sav';
string newvariablename (a5);
compute newvariablename = substr(CNTLNUMS,1,5);
sort cases by newvariablename;
save outfile = 'temporarydatafile.sav';
get file = 'districtdatafile.sav';
string newvariablename (a5);
```

```
compute newvariablename = substr(CNTLNUMD,1,5);
sort cases by newvariablename;
save outfile = 'temporarydistrictdatafile.sav';
match files file = 'temporarydatafile.sav'
        /table 'temporarydistrictdatafile'
        /by newvariablename;
save outfile = 'mergeddatafile.sav';
```

Sample Stata Syntax for Merging Data Files

Merging Restricted-use Data Files Using the School Control Number (CNTLNUMS)

When merging any of the school, principal, teacher, or school library media center data files together for a given school, the school's control number, CNTLNUMS, is used to merge data files. The Stata syntax is provided below. Notice that both data files being merged must be sorted by the school control number prior to performing the merge. Words in italics are meant to be replaced by meaningful file or variable names.

```
use dataset1
sort CNTLNUMS
save dataset1, replace
use dataset2
sort CNTLNUMS
save dataset2, replace
merge CNTLNUMS using dataset1
```

Merging the Restricted-use Public School District Data File with Other Public Sector Data Files

There are two ways to merge the public school district data file with other public sector data files. The first is with the district's control number (CNTLNUMD). This variable is included on the public school district data file as well as the public school data file. The sample syntax provided above is correct, except that the merging variable will be CNTLNUMD. However, since CNTLNUMD is not included on the principal, teacher, or school library media center data file, merging the public school district data file with these data files requires a different approach. Users will also need to use this method if the school did not respond to SASS.

The second method parses out the first five digits of the district's and the school's control number. The first five digits of CNTLNUMS and CNTLNUMD are identical, so users can create a new variable using a substring of these control numbers and merge the data files by the new variable name. The Stata syntax provided below illustrates how to merge the public school district data file with other data files using a substring. Please note that the data files being merged must be sorted by the variable listed in the "sort" statement prior to performing the merge. Users may need to increase memory before beginning the merge.

```
use districtfile
generate newvariablename = substr(CNTLNUMD,1,5)
sort newvariablename
save tempdistrictfile, replace
use school_or_principal_or_teacher_or_libraryfile
generate newvariablename = substr(CNTLNUMS,1,5)
sort newvariablename
```

        save *tempschool_or_principal_or_teacher_or_libraryfile*, replace
        merge *newvariablename* **using** *tempdistrictfile*
        drop if _merge= =2
        save *mergedfile*, **replace**

Unless specified, the default name of the merge variable created during the merging of files is _merge. The variable _merge identifies the various categories of data in a one-to-one merge. For example, if users merge the public school district ("using" data file) file onto the principal file ("master" data file):

    _merge= =1   observations from principal data file, no public school district data added (occurs
                               with district nonresponse)
    _merge= =2   observations from only public school district data file (e.g., district responded,
                               but there is no principal from that district)
    _merge= =3   observations from public school district and principal data files

By dropping the _merge= =2 observations, the merged data file will contain only principals, regardless of whether their district responded. No observations will remain when a district responded without a principal.

Chapter 11. User Notes and Cautions

The following notes cover cautions concerning change estimates with particular emphasis on the locale codes, population estimates for teachers that were produced by the 2004–05 Teacher Follow-up Survey (TFS) and the 2003–04 Schools and Staffing Survey (SASS), imputation of missing data for the created variable for percent minority enrollment (MINENR), and user notes and cautions for the 2003–04 SASS.

Users may also be interested in examining the crosswalk of variables contained in "Appendix K. Crosswalk of Items in 2004–05 TFS with 2000–01 TFS and 2003–04 SASS Teacher Questionnaire." This appendix has crosswalks for both TFS questionnaires.

Cautions Concerning Change Estimates

Care must be taken in estimating change over time in a TFS data element, because some of the measured change may not be attributable to a change in the educational system, but due to changes in the sampling frame, questionnaire item wording, or other changes. For example, the definitions of the locale codes based on the U.S. Census were revised in 2000 and again in 2003. Changes in how schools' locales are categorized over time may account for at least some changes that are noted from previous administrations. This impacts the urbanicity variables included in the data files, which are based on the 2000 Census definitions for locale codes.

The definition of locale codes changed between the 1999–2000 and 2003–04 administrations of SASS. To facilitate the transition, locale codes based on geographic concepts from both the 1990 and 2000 Decennial Census are included on the 2003–04 SASS data files. (SLOCP_99 uses the 1990 Census metropolitan areas, and SLOCP_03 and URBANS03 use the 2000 Census metropolitan areas.) The specific categories reported in the locale codes are based, respectively, upon the 1990 or 2000 definitions for central city, urban fringe of large or medium-sized central city, large or small town, and rural areas either inside a metropolitan area or outside a metropolitan area. The 1990 Decennial Census geographic areas were based upon countywide definitions of metropolitan or nonmetropolitan areas. By the 2000 Census, urban and rural classifications were based on a subcounty level.

In 2003, the Office of Management and Budget changed the geographic classifications, replacing "central city" with "principal city" and "Standardized Metropolitan Statistical Area" (SMSA) with "Core-based Statistical Area" (CBSA). However, these newer terms and locale codes could not be used in the 2003–04 SASS because the 2003 geographic classification of schools or school districts had not been completely implemented into the Common Core of Data (CCD) or the Private School Universe Survey (PSS), which serve as the sampling frames for SASS, by the time the 2003–04 SASS data were collected. Since then, the 2003–04 CCD and 2003–04 PSS have incorporated a new set of 12-level locale codes.

Only the urbanicity variables that utilize the 2000 definitions (SLOCP_03 and URBANS03) are included on the 2004–05 TFS data files. Therefore, caution should be taken when comparing urbanicity estimates from the 2004–05 TFS of the respondent's base-year school with previous administrations, because the locale codes are not based on the same definitions.

Estimates for Total Teachers

The total population estimate of teachers produced in TFS is slightly lower than that produced in SASS. The discrepancy is due to the fact that there were 99 respondents in SASS who were out of scope for TFS, because they had died or moved out of the country. These teachers were removed from the TFS sampling

frame. (See chapter 3 for more details on the sampling frame and chapter 8 for more details on weighting.)

Minority Enrollment

The variable measuring the number of minority students as reported by the SASS school (NMINST_S) is copied from data on the SASS school data files and placed on the TFS data files. The variable was imputed for 486 teachers on TFS whose schools did not respond to SASS.

There are three respondents in the TFS data files who have a value for the variable measuring the percentage of minority students in the SASS school (MINENR) in TFS but have missing data (-9) in the SASS data files. The values for MINENR were clerically imputed for the TFS data files. The missing data were imputed on TFS in order to produce the variable NMINST_S (number of minority students as reported by the SASS school) for TFS teachers whose corresponding base-year school did not respond to SASS.

User Notes and Cautions for SASS

Please see chapter 12 in *Documentation for the 2003–04 Schools and Staffing Survey* (Tourkin et al. 2007) for more details on user notes and cautions for data collected in the 2003–04 Schools and Staffing Survey (SASS). Notes and cautions were provided on the following:

- editing of the created variable for percent minority enrollment (MINENR);
- deletion of Title I data in Bureau of Indian Affairs (BIA) funded schools;
- data anomalies in created variables;
- the effect of missing data across files;
- locale codes used on the 2003–04 SASS;
- departmentalized and elementary enrichment teachers with no reported classes; and
- the existence of leading spaces on certain character variables in extracted SPSS files.

References

Efron, B. (1982). *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: SIAM.

Kaufman, S. (1992). Balanced Half-sample Replication with Aggregation Units. *1992 Proceedings of the Section on Survey Research Methods* (pp. 440–445). Alexandria, VA: American Statistical Association.

Kaufman, S. (1993). A Bootstrap Variance Estimator for the Schools and Staffing Survey. *1993 Proceedings of the Section on Survey Research Methods* (pp. 675–680). Alexandria, VA: American Statistical Association.

Kaufman, S. (1994). Properties of the Schools and Staffing Survey's Bootstrap Variance Estimator. *1994 Proceedings of the Section on Survey Research Methods* (pp. 1116–1121). Alexandria, VA: American Statistical Association.

Kaufman, S. (1998). A Bootstrap Variance Estimator for Systematic PPS Sampling. *1998 Proceedings of the Section on Survey Research Methods* (pp. 769–774). Alexandria, VA: American Statistical Association.

Kaufman, S. (2001). A New Model for Estimating the Variance under Systematic Sampling. *2001 Proceedings of the American Statistical Association*. Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association.

Kaufman, S. (2002). Using the Bootstrap in a Two-Stage Nested Complex Sample Design. *2002 Proceedings of the American Statistical Association* (pp. 1776–1781). Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association.

Kaufman, S. (2003). The Efficiency of the Bootstrap under a Locally Random Assumption for Systematic Samples. *2003 Proceedings of the American Statistical Association* (pp. 2097–2102). Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association.

Kaufman, S. (2004). Using the Bootstrap in a Two-Stage Sample Design When Some Second-Stage Strata Only Have One Unit Allocated. *2004 Proceedings of the American Statistical Association* (pp. 3766–3773). Survey Research Methods Section [CD-ROM]. Alexandria, VA: American Statistical Association.

Research Triangle Institute. (2001). *User Manual, Release 8.0*. Research Triangle Park, NC: Research Triangle Institute.

Seastrom, M.M. (2003). NCES Statistical Standards (NCES 2003-601). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved November 30, 2006, from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2003601.

Sitter, R.R. (1990). *Comparing Three Bootstrap Methods for Survey Data* (Tech. Rep. No. 152). Ottawa, Canada: Carleton University, Laboratory for Research in Statistics and Probability.

Tourkin, S.C., Warner, T., Parmer, R., Cole, C., Jackson, B., Zukerberg, A., Cox, S., and Soderborg, A. (2007). *Documentation for the 2003–04 Schools and Staffing Survey* (NCES 2007-337). U.S. Department of Education. Washington, DC: National Center for Education Statistics.