SPeNSE Final Report

Study of Personnel Needs in Special Education

Task 9.3-3

Final Report

Methods

Contract # ED00C00010

August 30, 2002

By

Elaine Carlson
Hyunshik Lee
Karen Schroll
Sheri Klein
Sharon Willig
Westat

Submitted to:

U.S. Department of Education
Office of Special Education Programs

Methods

The Study of Personnel Needs in Special Education (SPeNSE) provided nationally representative data on the workforce serving students with disabilities, including contextual information about the schools and districts in which these service providers worked. Designing and conducting a study like SPeNSE requires thousands of decisions, from the type of incentives offered to respondents to the method for nonresponse imputation. This chapter documents the study methods used in SPeNSE. It includes a description of major constructs and their measurement, sample design, data collection procedures, data cleaning, assigning sampling weights, addressing item nonresponse, analyzing the data, and using extant data. While the design work for SPeNSE was completed under a separate contract from that of the study's implementation, this chapter includes work completed under both contracts.

A number of terms used throughout this chapter require definition. In relation to individuals or groups of individuals serving students with disabilities, the term *quality* is used broadly. It includes qualifications, such as credentials, tested ability, and professionalism, as well as classroom teaching practices. For the workforce as a whole, as opposed to individual service providers, quality also includes demographic characteristics because of concern about pervasive differences in the demographics of students and their teachers. No quality scale or rubric was developed prior to the study so, in this document, quality is not used in an evaluative manner.

As used in this document, *service providers* and *personnel* include general and special education teachers (including preschool teachers), speech-language pathologists, and special education paraprofessionals who provide direct services to students. Personnel who provide services to other professionals, work exclusively with parents, or work in administration are excluded. The term *teachers* is narrower than *service providers* and includes only general and special education teachers (including preschool teachers).

Major Constructs and Their Measurement

SPeNSE relied on two primary data collection instruments: a survey of local special education administrators and a survey of service providers. The data were collected through computer-assisted telephone interviews (CATI) in 2000. The instruments were developed to describe the workforce and factors affecting workforce quality. In this section, we briefly describe the constructs explored in SPeNSE and the ways they were measured through the SPeNSE instruments. Complete copies of the instruments are available in Appendix B.

Classroom Teaching Practices

Teaching practices lie at the center of workforce quality, because interactions between service providers and their students most directly affect the outcome of interest, improved student achievement. In developing survey items, we drew upon the Council for Exceptional Children's (CEC) standards for entry into practice (CEC, 1998). Service providers were asked how skillful they are in those areas specified by CEC, such as planning effective lessons, managing instructional time, and working with parents. In addition to the skills drawn from the CEC standards, the study awards particular attention to five instructional areas: teaching reading, managing behavior, facilitating secondary transition, teaching English-language learners, and promoting inclusion. In each of these professional areas, service providers were asked the extent to which they use various best practices identified by experts in the field. For example, 12 instructional practices were listed for teaching reading, and respondents were asked, for each of the 12, whether they use that approach "never," "rarely," "sometimes," or "often."

The variables measuring classroom teaching practices were imperfect in that they relied on self-report, and it is not clear to what extent teachers accurately assessed their own use of best practices. Likewise, no independent psychometric work was done before administering the surveys to determine the extent to which use of these practices were associated with improvements in student performance. However, previous process-product research has shown that specific teaching practices are related to student achievement (Pressley, Wharton-McDonald, Allington, Block, & Morrow, 1998; Wenglinsky, 2000). For example, previous research consistently links practices, such as teaching phonics or phonemic skills, systematically teaching vocabulary, studying the style or structure of a text, having students summarize what they have read, and reading aloud with improvements in reading achievement (Pressley, Wharton-McDonald, Allington, Block, & Morrow, 1998; U.S. Department of Health and Human Services, 2000).

Credentials and Tested Ability

Previous research suggests that teacher credentials and tested ability are correlated with student achievement, at least for students without disabilities (Darling-Hammond, 1999; Ferguson, 1991; Goldhaber & Brewer, 2000; Greenwald, Hedges, & Laine, 1996; Hanushek, 1971). As such, SPeNSE service provider instruments included questions on participation in and performance on tests required for certification, certification for the position held, years of experience, and qualifications for working with diverse student populations.

Professionalism

The growing interest in the professionalization of teaching stems from the view that improving the status, training, and working conditions of teaching is central to addressing teacher quality. Theory holds that upgrading the status of teaching will lead to increased teacher commitment and motivation, which, in turn, will lead to improved teacher effectiveness and, finally, to improved student outcomes (Ingersoll, Alsalam, Quinn, & Bobbitt, 1997; Talbert & McLaughlin, 1994). SPeNSE service provider interviews included a number of items designed to measure professionalism. They included items on commitment to the profession; participation in professional activities, such as mentoring, membership in professional associations, reading professional journals; engaging in professional development beyond what is required by the district or State; and a self-efficacy scale. However, analysis of these variables did not provide evidence that a single latent professionalism construct exists. Rather, many of the variables appeared to function independently. A narrower construct, which Westat called *professional activities,* emerged through a factor analysis. It included reading professional journals, belonging to professional associations, and the frequency with which special educators were asked advice by their colleagues.

Demographic Representation

The increasingly diverse student population and the constancy in the characteristics of the teaching population (largely white, suburban, middle-class, and female) mean that students typically do not have opportunities to interact with or observe adults in school who are like themselves. Familiarity with different communication styles, cognitive orientations, and societal norms may enhance the instructional approaches educators use (Garcia & Malkin, 1993). Results from the SPeNSE service provider interviews provided a demographic profile of the workforce serving students with disabilities. Variables included sex, age, race, ethnicity, and disability. These were used to describe workforce demographics at the regional and national levels.

Working Conditions

Working conditions, such as work assignment, caseload, and administrative support may affect both personnel quality and staff turnover. If personnel are dissatisfied with the conditions in which they work, they may seek employment elsewhere. If wages increase, the quality of those entering the profession may improve. Many studies have supported the relationship between teacher retention and specific working conditions, such as student characteristics, paperwork requirements, job-related stress, class size, caseload, and perceived administrative support (Boe et al., 1997; George, George, Gersten, &

Grosenick, 1995; Metzke, 1988; Morvant et al., 1995; Singer, 1992; Singh & Billingsley, 1996). SPeNSE service provider interviews included a wide variety of items related to working conditions. These included school climate, age/grade level taught, disabilities of students, caseload, work responsibilities, and perceived level of support from administrators and colleagues.

State and Local Policies

State and local policies may also influence the quality of personnel employed and turnover rates. The National Association of State Boards of Education (NASBE) Study Group on Teacher Development, Supply and Demand (1998) argued that, "although hiring and placing teachers is a district and school responsibility, ensuring an adequate supply of teachers is fundamentally an issue of State policy that requires far-reaching, systemic reform to build a coherent system of teacher recruitment, development, and support (p. 4)." SPeNSE administrator interviews were used to collect data on local policies and procedures that might affect workforce turnover and quality. These included compensation policies, tenure policies, recruitment and retention practices, and professional development policies. Data on State personnel policies were drawn from extant data sources, such as the manual of the National Association of State Directors of Teacher Education and Certification (NASDTEC, 2000).

As part of the SPeNSE data collection, participating districts, intermediate education units (IEUs), and State schools were asked to submit salary schedules for 1997-98, 1998-99, and 1999-2000. Project staff compared the structure of the salary schedules to identify common variables that could be used for analysis. Items common to almost all salary schedules included base salary for teachers with a bachelor's degree, base salary for teachers with a master's degree, median salary for teachers with a master's degree, and maximum salary paid by the agency. These data were extracted from the salary schedules in participating districts, IEUs, and State schools. Salary variables were merged with the analytic database.

While absolute salaries may influence recruitment and retention, it is possible that salaries relative to other agencies in the State are even stronger predictors. As such, Westat also used cost-of-living-adjusted and unadjusted State average teacher salaries to create additional salary variables. For example, in addition to using median salary for a teacher with a master's degree as a predictor of staff turnover, analysts also used deviation from the State average as a measure of relative compensation.

Preservice Education

The knowledge and skills prospective educators acquire before taking their first job are critical to workforce quality and may also relate to attrition. Previous research suggests that the selectivity of the colleges that teachers attend and the amount of education coursework teachers complete are correlated with teacher performance and/or student achievement (Ehrenberg & Brewer, 1994; Ferguson & Womack, 1993). Recent studies in general education show that graduates of extended preservice programs are more satisfied with their preparation, are viewed by their colleagues as better prepared, are as effective with students as more experienced teachers, and are more likely than their traditionally prepared peers to stay in teaching (Darling-Hammond, as cited in Darling-Hammond & Ball, 1997). Furthermore, the number of applied teaching experiences in preservice education and graduate school experiences were both negatively associated with attrition for special educators (Mani, 1989; Metzke, 1988).

SPeNSE interviews with service providers collected extensive information on the characteristics of respondents' preservice programs, including type of certification program, weeks of student teaching, and overall perceived quality. In addition, respondents were asked where they completed their preservice preparation. Westat analysts used extant data about preservice programs in conjunction with data from the interviews to explore relationships between preservice preparation, teacher quality, and plans to remain in teaching.

Continuing Professional Development

Frechtling, Sharp, Carey, and Vaden-Kiernan (1995) characterized inservice professional development programs based on two general dimensions--focus and structure. The focus of professional development activities refers to the knowledge, skills, or attitudes professionals are expected to acquire through exposure to the professional development program. The structure of the professional development activities may be characterized by intensity, target population, and geographic scope. In SPeNSE interviews, service providers were asked the total hours of professional development in which they had engaged in the past year, as well as the specific form and focus of professional development activities. One set of interview items dealt specifically with induction activities in which the respondents participated.

Sample Design

The sample design for SPeNSE resulted in a large, nationally representative sample of personnel serving students with disabilities. To arrive at that sample, project staff used a two-phase sample design.

The first-phase sample was a nationally representative sample of local education agencies (LEAs), IEUs, and State schools for students with visual or hearing impairments. The second-phase sample was a representative sample of special education personnel from those selected LEAs, IEUs, and State schools.

First-Phase Sample

The first-phase sample of LEAs was a stratified simple random sample, which was drawn from the November 5, 1998, version of the Quality Education Data (QED). The LEA sample was stratified by geographic region and LEA size. The geographic regions correspond with those served by OSEP's six Regional Resource Centers as follows:

Region Code	Region Name	Member States
1	Northeast	CT, ME, MA, NH, NJ, NY, RI, VT
2	Mid-south	DC, DE, KY, MD, NC, VA, SC, TN, WV
3	Southeast	AL, AR, FL, GA, LA, MS, OK, TX
4	Great Lakes	IL, IN, IO, MI, MN, MO, OH, PA, WI
5	Mountain Plains	AZ, CO, KS, MT, NE, NM, ND, SD, UT, WY
6	Western	AK, CA, HI, ID, NV, OR, WA

LEA enrollment was used as the size measure for LEA size stratification. LEA enrollment was grouped into four categories: very large (50,000 or more), large (10,000-49,999), medium (2,500-9,999), and small (fewer than 2,500).[1]

The first-phase sample of IEUs was also selected from the QED. Agencies that did not employ staff who provide direct services to students with disabilities were deleted from the frame. However, this cleaning of the IEU frame was not perfect, and some ineligible IEUs were sampled. The IEU sample was stratified by geographic region.

All State schools serving students with visual or hearing impairments were included in the first-phase sample. The stratum population and sample sizes for first-phase sampling and the results of efforts to recruit the samples of LEAs, IEUs, and State schools for the study are presented in Tables 1, 2, and 3.

[1] These size strata are also referred to by numbers, 1, 2, 3, and 4, for the very large, large, medium, and small strata, respectively.

Table 1. LEA Population, Sample Sizes, and Recruitment Results

Very Large LEAs					
Region	Population Size	Sample Size	Target Number	Acceptances	Response Rate
1	2	2	2	1	50.0
2	18	18	14	8	44.4
3	30	30	24	16	53.3
4	7	7	6	4	57.1
5	8	8	6	2	25.0
6	11	11	9	9	81.8
Total	76	76	61	40	52.6
Large LEAs					
Region	Population Size	Sample Size	Target Number	Acceptances	Response Rate
1	44	9	5	4	44.4
2	117	20	12	12	60.0
3	172	31	18	14	45.2
4	141	22	15	9	40.9
5	70	10	7	5	50.0
6	182	45	20	14	31.1
Total	726	137	77	58	42.3
Medium LEAs					
Region	Population Size	Sample Size	Target Number	Acceptances	Response Rate
1	590	45	21	16	35.6
2	362	19	12	10	52.6
3	548	29	19	12	41.4
4	1,052	68	37	30	44.1
5	168	10	6	6	60.0
6	403	21	14	9	42.9
Total	3,123	192	109	83	43.2
Small LEAs					
Region	Population Size	Sample Size	Target Number	Acceptances	Response Rate
1	1,730	37	23	13	35.1
2	247	5	3	2	40.0
3	1,745	45	23	11	24.4
4	3,329	83	45	36	43.4
5	1,829	42	24	19	45.2
6	1,046	27	14	9	33.3
Total	9,926	239	132	90	37.7
Grand Total	**13,851**	**644**	**379**	**271**	**42.1**

Table 2. IEU Population, Sample Sizes, and Recruitment Results

Region	Population Size	Sample Size	Target Number	Eligible	Acceptances	Response Rate
1	113	9	4	6	3	50.0
2[a]	26	26	2	1	1	100.0
3	191	23	7	14	4	28.6
4	468	36	17	34	22	64.7
5	226	18	9	16	11	68.8
6	96	7	4	7	5	71.4
Total	**1,120**	**119**	**43**	**78**	**46**	**59.0**

[a]Note: All IEU's were selected due to the unusually low eligibility rate in region 2.

Table 3. State School Population, Sample Sizes, and Recruitment Results

Region	Population Size	Sample Size	Target Number	Acceptances	Response Rate
1	5	5	4	3	60.0
2	16	16	11	10	62.5
3	17	17	12	14	82.4
4	16	16	11	12	75.0
5	12	12	8	8	66.7
6	10	10	7	6	60.0
Total	**76**	**76**	**53**	**53**	**69.7**

The participation (response) rates for the LEA, IEU, and State school samples are 42.1, 59.0, and 69.7 percent, respectively. This resulted in an overall participation rate of 47.7 percent.

The second-phase sample was a stratified simple random sample of service providers from rosters of personnel that were obtained from 370 participating LEAs, IEUs, and State schools.[2,3] The total number of providers in the sampling frame was 315,047.

In early spring 2000, Westat sent letters to administrators in participating districts with a list of service providers who had been sampled for the study. Administrators were asked to verify that the sampled service providers still worked in the district, intermediate unit, or State school; that their job was

[2] After service provider samples were selected, one district (size 1, region 5) pulled out of the study before the provider samples were loaded into the CATI system. The sample of service providers for size 1, region 5 was supplemented by selecting additional service providers from other districts (see the Memo of May 15, 2000, by Hyunshik Lee). A second district dropped out after CATI loading, and there were two provider interviews completed at that time. Therefore, the second district is included as successfully recruited, but the first is treated as a recruitment nonrespondent.

[3] The rosters and sample from IEUs included only special education teachers. The roster and sample from State schools included only special education teachers who serve primarily students with visual or hearing impairments.

accurately coded; and that names, addresses, and phone numbers were accurate. Service providers who had retired prior to the 1999-2000 school year or were no longer working within the sampled school district in 1999-2000 were made ineligible unless the district provided the name of their replacement, in which case the replacement became part of the sample. Administrators were also asked to indicate if any of the potential respondents required accommodation (Braille letters, Spanish-speaking interviewer, or a sign language interpreter onsite to conduct the interview) to participate. Updated lists of service providers were faxed back to Westat. Westat received verification information from 352 of the 370 education agencies that had agreed to participate.

Among the 370 participating agencies, 47 LEAs failed to provide rosters for general education teachers, and 11 of them did not provide rosters of paraprofessionals (see Table 4). The providers from the missing rosters were treated as nonsampled providers. This approach was preferable to treating the agencies as nonrespondents for sampling general education teachers or paraprofessionals for two reasons. First, it allowed us to use one set of agency weights for all personnel types. Second, less variable weights were produced. However, this approach required us to estimate the size of the missing rosters.

This estimation was done using the information on the total number of teachers at the agency level available on QED and a simple ratio model that related the total number of teachers and the number of general education teachers or paraprofessionals. It was estimated that 90.9 percent of teachers were general education teachers, and the ratio of the total teacher body to the special education paraprofessionals was 9.1 percent. Using these estimated parameters of the ratio models, the size of missing rosters was estimated as shown in Table 4.

The basic principle for the second-phase sample allocation was to minimize the difference among the sampling weights, knowing the first-phase sampling weights (based on the realized agency sample sizes) and the number of service providers of each type. Using this principle, sample allocation was done separately for LEAs, IEUs, and State schools. Seven types of personnel were sampled:

- Special education teachers who served primarily children with disabilities ages 3-5,
- Special education teachers who served primarily students with visual or hearing impairments,
- Special education teachers who served primarily students with emotional disturbance,
- Special education teachers who were not included in the previous three categories,
- Speech-language pathologists,
- Special education paraprofessionals, and

- General education teachers

Table 4. Number of LEAs with Missing Rosters and Estimated Number of Service Providers

Size	Region	General Ed. Teacher		Paraprofessional	
		LEA	Estimated	LEA	Estimated
1	2	5	26,495	-	0
1	4	3	34,891	-	0
1	5	1	2,460	-	0
1	6	4	20,740	1	2,722
2	2	5	3,251	-	0
2	3	1	896	1	131
2	4	4	4,703	-	0
2	5	1	1,245	-	0
2	6	3	1,209	1	3
3	1	5	771	2	49
3	2	3	789	2	46
3	3	1	305	-	0
3	4	4	914	1	29
3	6	3	774	-	0
4	1	2	152	-	0
4	3	1	41	-	0
4	4	1	53	3	10
Total		47	99,689	11	2,990

Second-Phase Sample

The allocation formula for personnel type stratum *i* within a first-phase design stratum is given as:

$$n_i \propto W_i N_i$$

where

n_i = the allocated second-phase sample size,
W_i = the first-phase sampling weight, and
N_i = the number of service providers.

Thus, the sample size n_i is determined by:

$$n_i = T \frac{W_i N_i}{\sum W_j N_j}$$

where T is the total sample size to be allocated.

In many cases, there were not enough providers to select the needed sample, so all available service providers were selected. This necessitated cycling through the allocation process a few times. Allocation of the sample to equalize the sampling weights was feasible only for general education teachers and paraprofessionals. Although these personnel strata suffered from missing rosters, there were still enough providers in the frame.

Conversely, there were several stratum cells that had no service providers and thus, no provider sample. For the cases where the number of providers were estimated due to missing rosters, the estimated N_i's were used in the allocation. This, in effect, made sampled providers represent those in the districts that were missing rosters; these service providers had no chance to be selected. To address bias that may have been introduced in this way, post-stratification (described in detail later) was used. The sample allocation results are presented in Appendix C.

The participating district in size 1/region 1 allowed only 100 providers in the sample, even though 179 were initially selected. Therefore, its sample had to be trimmed; the new sample sizes are reflected in Appendix C.

Data Collection Procedures

In late spring 2000, Westat sent letters to sampled administrators and service providers describing the study and asking them to participate by calling Westat's Telephone Research Center (TRC) to schedule an interview. Included in the letter was a postage paid postcard that potential respondents could use to schedule appointments. The letter to service providers also informed potential respondents that they would be entered into a drawing for one of several gift certificates (a $2,000 gift certificate to Circuit City, 10 $250 gift certificates to Amazon.com, and 50 $10 gift certificates to Starbucks Coffee). If prize winners did not have access to these retailers, alternate prizes of equal value were provided. Follow-up postcards and letters were mailed 4 weeks after the initial mailing, and TRC staff called potential respondents at work to try to schedule appointments or complete interviews.

Letters to administrators included a pre-survey. The pre-surveys were designed to assist administrators in organizing information they would need to complete the telephone interview. Separate versions of the pre-survey were sent to administrators in LEAs, IEUs, and State schools for students with visual or hearing impairments. Copies of the pre-surveys are included in Appendix D.

SPeNSE interviews were conducted from May 2000 through November 2000. For those interviews conducted in fall 2000, we changed the survey items to past tense so the administrators and providers would base their answers on the previous school year (1999-2000). Telephone interviews were conducted by professional interviewers at Westat's Frederick, Maryland, and Sarasota, Florida, locations. Seventy interviewers, in all, worked on the study. They received 2 days initial training and follow-up instruction, as needed. Interviews were monitored on site as well as through remote access.

Because of the length of the interviews and the difficulties teachers found in allocating time during the work day to complete them, respondents were permitted to complete interviews in segments. Table 5 shows the average length of interviews for each type of service provider and for administrators.

Table 5. Mean Length of SPeNSE Interviews in Minutes, by Type of Personnel

Special education teachers who serve primarily children with disabilities ages 3-5	53.8
Special education teachers who serve primarily students with visual or hearing impairments	59.0
Special education teachers who serve primarily students with emotional disturbance	52.7
Special education teachers who are not included in the previous three categories	55.4
Speech-language pathologists	55.5
Special education paraprofessionals	27.7
General education teachers	49.4
Administrators	29.8

As part of the process of preparing the analytic data set for the service providers, Westat staff identified and updated the cases in which an interview was partially completed. They singled out cases where: a) the case was not coded Complete or Ineligible and b) the case was broken off in section D2 of the survey or later; 155 cases fit this criteria. They were treated as completed interviews. Partially completed interviews for 1,663 respondents were treated as nonresponses.

The first section of the service provider interview screened respondents to ensure that they were eligible to participate and that they were employed in the capacity indicated by their administrator. Individuals who were deemed ineligible were thanked for their time and the interview was terminated. Eligible individuals who were misclassified were reclassified within the CATI system, and the interview continued using the corrected personnel type.

Despite efforts to develop a clean sampling frame, during the verification and interview processes, many sampled service providers were deemed ineligible because they had not worked in the 1999-2000 school year, were not employed in one of the specified job-types, no longer worked in the sampled education agency, died or were on long-term leave. Many of those not employed in one of the specified job types were school psychologists, social workers, and counselors. In all, 956 potential respondents were found ineligible through verification, the screener, or, in a few cases, based on responses to survey items or comments made during the interview that were recorded by the interviewer (e.g., no hours per week teaching, do not provide direct services to students).

The low rate of response among teachers of students with visual and hearing impairments was due to the itinerant nature of these teachers' jobs. They often work in five or more schools each week and have no central location at which they can be reached by telephone.

Table 6 shows the total number of eligible service providers and administrators sampled and the number and percentage of respondents, by type of personnel.

Table 6. Number of Eligible Personnel Sampled and Response Rates, by Type of Service Provider

Type of Service Provider	Sample Size[4]	Response Rate	
		Number	%
Special education teachers who serve primarily children with disabilities ages 3-5	1,171	881	75.2
Special education teachers who serve primarily students with visual or hearing impairments	1,679	1,054	62.8
Special education teachers who serve primarily students with emotional disturbance	1,190	859	72.2
Special education teachers who are not included in the previous three categories	3,688	2,633	71.4
Speech-language pathologists	1,261	885	70.2
Special education paraprofessionals	1,193	888	74.4
General education teachers	1,356	861	63.5
Total Service Providers	11,538	8,061	69.9
Administrators	370[5]	358	96.7

[4] Excludes those sampled individuals found ineligible in the verification process or the screening portion of the interview, individuals who died or became incapacitated between the time when the sampling frame was developed and the data collection period.

Using Extant Sources of Data

SPeNSE drew data from a variety of sources. These included CATI interviews with local administrators, CATI interviews with service providers, and extant data. Extant data sources included the QED; the National Clearinghouse for Professions in Special Education (NCPSE) database on preservice programs; and State policies summarized by the National Association of State Directors of Teacher Education and Certification (NASDTEC). This section describes uses of data from each of the extant data sources.

Data from the QED

When the SPeNSE sample of LEAs and IEUs was selected from the QED, a number of background variables on the sampled agencies were also extracted. These include Orshansky poverty index and metropolitan status. These were used as explanatory variables as the SPeNSE staff explored factors affecting the workforce serving students with disabilities.

Data from the National Clearinghouse for Professions in Special Education (NCPSE) Database on Preservice Programs

NCPSE at CEC maintains a database on preservice programs in the country that certify special education teachers. The database includes variables such as area of program emphasis; degrees awarded; number of special education faculty; availability of dual, blended, or integrated programs; and whether the program is NCATE certified. In the SPeNSE interview, teachers were asked where they completed their preservice education. By matching information from the NCPSE database with information on the school at which respondents completed their preservice education, researchers explored the relationships among preservice training and personnel qualifications. Because preservice programs change over time, caution was used in interpreting results for individuals who completed their programs years ago.

Before using the NCPSE database on preservice programs, Westat completed a number of activities to verify the completeness and accuracy of the data. First, Westat analysts reviewed the procedures used by CEC for collecting and verifying the data. Second, Westat analyzed the completeness

---

[5] This is the total number of districts, IEUs, and State schools that agreed to participate from 824 LEAs, IEUs, and State schools that were invited. Participation involved submitting a roster of personnel from which the sample of service providers could be selected and agreeing to have an administrator complete an interview.

of the data. Third, Westat analysts extracted a small sample of records (i.e., nine) and verified the information using university web sites or telephone interviews with university administrators. With minor exceptions, the data were found to be valid and reliable.

Data from the NASDTEC on State Policies

NASDTEC published tables that summarize a broad range of State policies related to teacher certification and licensure. From this publication, Westat extracted data on State tests required for certification or licensure. Variables were added to the SPeNSE data set to indicate which States required a test of teaching performance, a basic skills test, a test of teaching knowledge, and a test of subject knowledge.

Data Cleaning

Many edit checks were built into the CATI instrument, for example, ranges of acceptable responses and checks for internal consistency. Nonetheless, extensive efforts were made to clean the data further, after data collection was completed. In particular, extensive data cleaning was conducted to ensure that responses were internally consistent. This involved a process of checking frequencies to every item to ensure that skip patterns worked correctly and respondents had a valid response for every item on the survey.

As part of that effort, we changed the reported teacher-type for 20 individuals. Nineteen teachers were coded as serving primarily students with emotional disturbance or behavioral disorders but had no students whose primary disability was emotional disturbance, autism, developmental delay, other health impairments (which included ADHD), or multiple disabilities. Administrators originally coded each of the respondents as another type of special education teacher. Researchers called a sample of these respondents to clarify the inconsistent responses and, in three of three cases, the respondents indicated that they did not teach students with emotional disturbance and felt it must have been a data entry error. In all 19 cases, the respondents were recoded as *other special education teacher*. A similar situation arose with one respondent who was classified as a teacher who served primarily students with visual or hearing impairments but who served exclusively students whose primary disabilities were learning disabilities. This individual was recoded as *other special education teacher*. Because the surveys were identical for these different types of teachers, no recoding of responses was required.

Open-ended responses to *other, specify* questions were examined, and frequencies were generated for those that were most common. Variables created as a result of the *other, specify* analyses begin with the letter O. Because those items were not asked directly of all respondents, they were not included in standard descriptive analyses. However, they were included in written summaries of the data. A detailed description of the ways in which these open-ended responses were coded is included in Appendix E.

Minor changes to the instruments were necessary after data collection began. Due to its complexity and length of administration, question SC6 in the provider instrument was removed from the survey at the beginning of data collection. All variables related to this question were removed from the database.

The skip pattern for question SC4 item g was programmed incorrectly until August 2000. All providers serving secondary-aged students should have been asked the question. However, because of a programming error, only providers who exclusively served secondary-aged students received the question; therefore, providers who served mixed-aged populations (i.e., both elementary and secondary-aged students) did not receive the question. The variable for this item was suppressed in the analytic data set.

After data collection ended, we identified an error in the skip pattern for question SD6. Providers serving students with mental retardation or multiple disabilities were to be asked item e on SD6. However, the CATI was programmed so that only providers who served both students with mental retardation and students with multiple disabilities were asked the item. The same problem occurred on items g, h, and i for SD6. Providers serving students with other health impairments or orthopedic impairments were to be asked items g, h, and i. However, the CATI was programmed so that only providers serving students with both orthopedic impairments and other health impairments were asked the items. Consequently, the cell size for those sub-items on SD6 is small, and results should be used with caution.

Assigning Sampling Weights

The weighting was done in phases reflecting the two-phase sample design. The weighting for the sample of LEAs, IEUs, and State schools was based on the recruitment results, with minor adjustments. First, the recruited agencies were weighted by adjusting the base weights (inverse selection probability) for nonparticipating agencies. These weights were later used for developing the service provider weights.

Second, the same weights were further adjusted for (12) nonresponding administrators who originally agreed to participate but did not complete administrator interviews. This section provides detail on that process.

The LEA participation rate at the recruitment stage was quite low; the overall rate was 42 percent. Moreover, the realized sample size reached only 72 percent of the target even with supplemental sampling. The IEU participation rate of 44 percent was also low, but the number of participants exceeded the target number when an extra sample was used to supplement the original sample. The originally selected LEA and IEU samples were supplemented during the recruitment process in an effort to reach the target number of participating agencies because of much lower response rates than anticipated (see the Final Report of the SPeNSE Design Work, Deliverable 12D, August 30, 1999). Only the State school sample met the expected participation rate of 70 percent and the target sample size.

We knew from the outset that the IEU sampling frame included some ineligible agencies. An eligibility rate of 70 percent was expected; the actual eligibility rate in the sample was 66 percent. However, the eligibility rate was not uniform across regions, ranging from 4 to 100 percent. In particular, a very low rate of eligibility in region 2 necessitated selecting all IEUs in the region during supplemental sampling, and we found that there was only one eligible unit from the whole region. Region 3 also had a low eligibility rate (about 60 percent) and needed a larger supplemental sample to reach the target number. These regions had a relatively small number of IEUs but contributed a disproportionately large number of ineligible IEUs, which resulted in a lower than expected eligibility rate in the sample. However, regions with a large number of IEUs had a much higher eligibility rate, and thus, the weighted eligibility rate was 83 percent, much higher than expected. The estimated eligible IEU population size was 931 out of 1,120 IEUs in the original sampling frame.

The response pattern was analyzed by an SPSS software, CHAID (Chi-squared Automatic Interaction Detector, SPSS, 1993), and logistic regression. The analysis indicated that the first-phase nonresponse adjustment weighting could be successfully carried out within each design stratum and thus, nonresponse weight adjustment was done within each stratum. There were originally 36 strata altogether (24 size-region cells in the LEA sample, 6 regions in each of the IEU and State school samples), but the very large size group in Region 1 was collapsed into the very large size group in Region 2 to facilitate variance estimation. This reduced the number of strata to 35, and they can be serially numbered from 1 to 35. Let w_h be the nonresponse adjusted weight for the units in h-th stratum for $h = 1, ..., 35$, and then it

is defined by $w_h = N_h / n_h$, where N_h is the stratum population size and n_h is the stratum respondent size (this should be distinguished from the original sample size).[6] For the IEU sample, the N_h were estimated. These first-phase weights were then used to calculate the administrator sample weights and service provider sample weights as explained below.

Weighting the Administrator Sample

Although they agreed to participate in the study at the recruitment stage and provided rosters for their agencies, a small number of administrators (12) failed to complete interviews and became nonrespondents for the administrator survey. The administrator response rate of 96.7 percent given in Table 6 is very high, but it is out of the 370 agencies that originally agreed to participate, not of the fielded sample of agencies. The overall administrator response rate is 46.2 percent. These administrator nonresponses were simply weight-adjusted within the design strata.

For variance estimation using the jackknife method, replicate weights were computed for the JKn method, which is suitable for a stratified simple random sample such as the administrator sample (Westat, 2000). The very large agencies in region 1 had only one respondent, and thus, the stratum was collapsed with the same size group in region 2 for the JKn replication formation. The JKn method creates one replicate for each sample unit, resulting in 358 replicates corresponding to the number of respondent administrators.

Weighting the Service Provider Sample

The service provider sample was selected in the second-phase sampling process, where the roster pool was stratified by personnel type within each first-phase stratum and a simple random sample of providers was selected from each personnel type stratum. However, as mentioned earlier, second-phase sampling was marred by the following sampling irregularities:

1) Out of 271 LEAs, 47 did not provide rosters of general education teachers as shown in Table 2;
2) Eleven LEAs out of the 47 LEAs did not give rosters of paraprofessionals;

[6] The stratum respondent size is given in the *acceptances* column in Tables 1, 2, and 3.

3) Two very large LEAs were selected with certainty from region 1 (Northeast) but only one participated.

For 1) and 2), the providers in missing rosters were treated as if they were randomly excluded from the sample. For 3), the two size 1 strata were collapsed into one to facilitate variance estimation. This collapsing reduced the total number of the first-phase strata to 35, which constituted the variance strata for the jackknife, which is described in more detail later in this chapter.

The term *cluster* is used to mean an agency-level unit of LEAs, IEUs, or State schools, which were combined to form a variance unit for the JKn variance estimator (see WesVar manual, 2000). The JKn method allows as many replicates to form as the number of variance units within each variance stratum. Combining of agencies was done only in the medium and small size groups and for the IEU and State school samples, with two goals in mind. First, small agencies were grouped together to get larger units. Second, it reduced the number of replicates for the JKn jackknife variance estimator. However, the numbers of the combined units that served as variance units are reasonably large to provide reasonable degrees of freedom for analyzing the data at subgroup level such as size groups or regions as shown in table 7. Note that the degrees of freedom in the jackknife variance estimation is determined by the number of replicates, and for the JKn method it is equal to the number of variance units.

Table 7. Number of Variance Units by Size Group and by Region

Size Group	Number of Variance Units	Region	Number of Variance Units
1	40	1	22
2	58	2	34
3	42	3	51
4	31	4	50
IEU	24	5	25
State School	26	6	39
Total	221	Total	221

The number of variance units originally created was 221 but one of them had no provider sample and, thus, dropped out. The variance units are called clusters in the following discussion.

We need some notation to discuss the weighting procedure in detail:

N_h: population size (i.e., the number of clusters) of stratum h, h = 1, 2, 3, …, 35;

n_h: cluster sample size for stratum h;

U_h: set of all providers in n_h sampled clusters from stratum h;

f_h: first-phase sampling fraction for stratum h, i.e., $f_h = n_h / N_h$;

u_h: provider sample selected from U_h;

w_{hi}: first-phase weight of provider i in stratum h;

y_{hi}: value of survey variable y for provider i in stratum h.

The stratum cluster sample size n_h is the realized sample size, and w_{hi} is defined based on this realized sample size as follows:

$$w_{hi} = \frac{N_h}{n_h}. \tag{0.1}$$

This definition is justified because nonresponse adjustment of the first-phase sample was done within each stratum. If there were no second-phase sampling of providers and all providers responded, then the total Y of the y-variable would be estimated by

$$\hat{Y}_1 = \sum_{h=1}^{35} \sum_{i \in U_h} w_{hi} y_{hi}. \tag{0.2}$$

Ignoring the finite population correction (fpc) $(1 - f_h)$, the jackknife variance estimator for this estimate would be

$$v_J(\hat{Y}_1) = \sum_{h=1}^{35} \left(\frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (\hat{Y}_1^{(hj)} - \hat{Y}_1)^2 \right). \tag{0.3}$$

where $\hat{Y}_1^{(hj)}$ is the jackknife replicate estimator of Y calculated from replicate (hj) with cluster (hj) deleted. It is defined as

$$\hat{Y}_1^{(hj)} = \sum_{h'=1}^{35} \left(\sum_{i \in U_{h'}} w_{h'i}^{(hj)} y_{h'i} \right), \tag{0.4}$$

where

$$w_{h'i}^{(hj)} = \begin{cases} 0 & \text{if } h' = h \text{ and } i \text{ is in cluster } hj \\ N_h/(n_h - 1) & \text{if } h' = h \text{ and } i \text{ is not in cluster } hj \\ N_h/n_h & \text{if } h' \neq h \end{cases} \tag{0.5}$$

for $j = 1, 2, ..., n_h$, and $h, h' = 1, 2, ..., H$. There are 221 replicates.

If the fpc were not ignored, the jackknife variance estimator would be

$$v_J(\hat{Y}_1) = \sum_{h=1}^{35} \left(\frac{(1 - f_h)(n_h - 1)}{n_h} \sum_{j=1}^{n_h} (\hat{Y}_1^{(hj)} - \hat{Y}_1)^2 \right). \tag{0.6}$$

Later, we will further discuss the implication of incorporating the fpc as in (0.6) for multiphase sampling.

These formulae (0.1 through 0.6) cannot be used for SPeNSE because there was actually second-phase sampling, and there was a considerable amount of nonresponse at the second-phase as well. However, the formulae provide a structure by which variance estimation can be done with appropriate weight modification to take into account of subsequent sampling and other weighting adjustments. To discuss this weight modification for second-phase sampling, we need additional notation.

G: number of second-phase (personnel type) strata in first-phase stratum h;

A_{hg}: set of providers in second-phase stratum g in first-phase stratum h before second-phase sampling;

a_{hg}: second-phase sample of providers from A_{hg};

M_{hg} : size of A_{hg} ;

m_{hg} : size of a_{hg} ;

$M_{hg}^{(hj)}$: size of $A_{hg}^{(hj)}$ after removing cluster (*hj*);

$m_{hg}^{(hj)}$: size of $a_{hg}^{(hj)}$ after removing cluster (*hj*);

To estimate *Y*, two estimators are available in the literature. One is the double expansion estimator (DEE) and the other is the reweighted estimator (REE) (Kim, Navarro, and Fuller, 2000; Kott and Stukel, 1997). The DEE is unbiased, but unless m_{hg} are small, it is less efficient than the REE even though the REE, as a type of ratio estimator, is biased. For the SPeNSE provider sample, however, the two estimators are identical. If we use the REE form, however, it is given by

$$\hat{Y}_2 = \sum_{h=1}^{35}\sum_{g=1}^{G}\left(\sum_{i\in a_{hg}} w_{hi}\frac{\sum_{i\in a_{hg}} w_{hi}y_{hi}}{\sum_{i\in a_{hg}} w_{hi}}\right) = \sum_{h=1}^{35}\sum_{g=1}^{G}\left(\sum_{i\in a_{hg}} w_{hi}^{*}y_{hi}\right) \tag{0.7}$$

where for $i \in a_{hg}$,

$$w_{hi}^{*} = \frac{\sum_{k\in A_{hg}} w_{hk}}{\sum_{k\in a_{hg}} w_{hk}} w_{hi} = \frac{M_{hg}}{m_{hg}}\frac{N_h}{n_h} \tag{0.8}$$

is the modified weight. The simplified form in (0.8) is also the DEE weight.

The jackknife variance estimator is given by

$$v_J(\hat{Y}_2) = \sum_{h=1}^{35}\left(\frac{n_h - 1}{n_h}\sum_{j=1}^{n_h}(\hat{Y}_2^{(hj)} - \hat{Y}_2)^2\right) \tag{0.9}$$

where $\hat{Y}_2^{(hj)}$ is the replicate estimator for replicate (*hj*) defined analogously by (0.7) with replicate weights given by (0.5) and is written as

$$\hat{Y}_2^{(hj)} = \sum_{h=1}^{35} \sum_{g=1}^{G} \left(\sum_{i \in a_{hg}} w_{hi}^{(hj)} \frac{\sum_{i \in a_{hg}} w_{hi}^{(hj)} y_{hi}}{\sum_{i \in a_{hg}} w_{hi}^{(hj)}} \right) \qquad (0.10)$$

$$= \sum_{h=1}^{35} \sum_{g=1}^{G} \left(\sum_{i \in a_{hg}} w_{hi}^{*(hj)} y_{hi} \right).$$

Note that $w_{hi}^{*(hj)}$ is obtained from (0.8) by replacing w_{hi} by $w_{hi}^{(hj)}$ and given as

$$w_{hi}^{*(hj)} = \begin{cases} 0 \text{ if } h' = h \text{ and } i \text{ is in cluster } (hj) \\ (M_{hg}^{(hj)} N_h) \Big/ \left\{ m_{hg}^{(hj)} (n_h - 1) \right\} \text{ if } h' = h \text{ and } i \text{ is not in cluster } (hj) \\ (M_{hg} N_{h'}) / (m_{hg} n_{h'}) \text{ if } h' \neq h \end{cases} \qquad (0.11)$$

for $i \in a_{hg}$, $j = 1, 2, ..., n_h$, and $h, h' = 1, 2, ..., H$.

Unit nonresponses at the second-phase were weight-adjusted within weighting classes, which were formed based on the results of CHAID analysis. CHAID analysis was carried out for each personnel type within each sample type (LEA, IEU, and State school), using different auxiliary variables depending on the sample type. For the LEA sample, size group and region variables were forced to be the first and second level predictors regardless of their prediction power of nonresponse probability, which amounted to create weighting classes within first-phase strata. Other auxiliary variables used in the CHAID analysis, which were district-level variables obtained from the January 2000 QED, included metropolitan status, Orshansky poverty index, education climate index, and percentage of white population; these variables were each collapsed into two or three categories.

For the IEU sample, only the region and metropolitan status variables were used, while only region was used for the State school sample. Not all auxiliary variables used for the LEA sample were available for the IEU and State school samples. Furthermore, the sample base was too thin to use refined classes.

The variables used to form the weighting classes are summarized in the following table:

Variable	Definition	Sample Type
Region	5 OSEP Regions (design variable but Regions 1 and 2 were collapsed)	LEA, IEU, State school
Size	4 categorized district sizes (design variable)	LEA
Metro status	3 categories: Urban, Suburban, Rural	LEA, IEU
Percent white	Percentage of white population grouped into two categories 1 – if white percentage is ≤ 54; 2 – otherwise.	LEA
Education climate index (ECI)	5 QED indices were grouped into three: 1 - ECI = 0, 1 or 2; 2 – ECI = 3; 3 – ECI = 4 or 5.	LEA
Orshansky poverty index	The index is categorized into two groups: 1 - Orshansky index ≤ 24; 2 – Otherwise.	LEA

The boundary points to create categories for continuous variables such as "percent white" and "Orshansky poverty index" were determined to get equal size categories (i.e., they are median points).

Some of the weighting classes obtained from the CHAID analysis were very small, which could lead to unstable adjustment factors. For this reason, a minimum of 20 providers was imposed upon the class size and a maximum of 2.0 upon the adjustment factor. In order to satisfy these requirements, it was necessary to collapse deficient cells with other adjacent cells. For a small number of cases, the latter rule was overruled if the adjustment factor was not too excessive (say < 3) and further collapsing was not desirable. At the end, 126 final weighting classes were created.

The nonresponse realization at the second-phase was treated as the third-phase sampling for the purposes of weighting and variance estimation by the jackknife method. More pieces of notation are needed to formalize this discussion. Let the number of weighting classes be denoted by D and let

R_d : set of sample providers in weighting class d;

r_d : set of respondent providers in R_d ;

for $d = 1, 2, ..., D$. The nonresponse-adjusted estimator in an REE form can be written as

$$\hat{Y}_3 = \sum_{d=1}^{D} \left(\sum_{i \in R_d} w_{di}^{*} \frac{\sum_{i \in r_d} w_{di}^{*} y_{di}}{\sum_{i \in r_d} w_{di}^{*}} \right) = \sum_{d=1}^{D} \left(\sum_{i \in r_d} w_{di}^{**} y_{di} \right) \tag{0.12}$$

where w_{di}^{*} is defined in (0.8) with i in first-phase stratum h and

$$w_{di}^{**} = \frac{\sum_{k \in R_d} w_{dk}^{*}}{\sum_{k \in r_d} w_{dk}^{*}} w_{di}^{*} \tag{0.13}$$

is the nonresponse-adjusted weight.

The jackknife variance estimator of this estimator is given by

$$v_J(\hat{Y}_3) = \sum_{h=1}^{35} \left(\frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (\hat{Y}_3^{(hj)} - \hat{Y}_3)^2 \right) \tag{0.14}$$

where $\hat{Y}_3^{(hj)}$ is defined analogously by (0.12) with replicate weights given by (0.11) and is written as

$$\hat{Y}_3 = \sum_{d=1}^{D} \left(\sum_{i \in R_d} w_{di}^{*(hj)} \frac{\sum_{i \in r_d} w_{di}^{*(hj)} y_{di}}{\sum_{i \in r_d} w_{di}^{*(hj)}} \right) = \sum_{d=1}^{D} \left(\sum_{i \in r_d} w_{di}^{**(hj)} y_{di} \right) \tag{0.15}$$

Note that $w_{di}^{**(hj)}$ is obtained from (0.13) by replacing w_{di}^{*} by $w_{di}^{*(hj)}$.

The weights given by (0.13) were further modified by post-stratification ratio adjustment using a district level QED variable, the number of teachers, to obtain the post-stratum control totals. Since this auxiliary variable was not available for State Schools, no post-stratification was done for the State school sample. For the IEU sample, post-stratification was not performed either because, although the variable was available, the quality of the data were not good. Post-strata were defined by the cross-classification of six regions, three classes of education climate index, and three metropolitan statuses. The control totals were compiled from the January 2000 version of QED.

To introduce the post-stratified estimator, let

K : number of post-strata;

P_k : set of provider population in post-stratum k;

p_k : set of sampled providers in post-stratum k;

x_{ki} : auxiliary variable (counting variable, which is always 1) to be used for post-stratification ratio adjustment;

Then the post-stratified estimator is given by

$$\hat{Y}_4 = \sum_{k=1}^{K} \left(\sum_{i \in P_k} x_{ki} \frac{\sum_{i \in p_k} w_{ki}^{**} y_{ki}}{\sum_{i \in p_k} w_{ki}^{**} x_{ki}} \right) = \sum_{k=1}^{K} \left(\sum_{i \in p_k} w_{ki}^{***} y_{ki} \right) \tag{0.16}$$

where

$$w_{ki}^{***} = \frac{\sum_{i \in P_k} x_{ki}}{\sum_{i \in p_k} w_{ki}^{**} x_{ki}} w_{ki}^{**} . \tag{0.17}$$

The jackknife variance estimator for $\hat{Y}_4$ is then given by

$$v_J(\hat{Y}_4) = \sum_{h=1}^{35} \left(\frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (\hat{Y}_4^{(hj)} - \hat{Y}_4)^2 \right) \qquad (0.18)$$

where

$$\hat{Y}_4 = \sum_{k=1}^{K} \left(\sum_{i \in p_k} w_{ki}^{***(hj)} y_{ki} \right) \qquad (0.19)$$

and

$$w_{ki}^{***(hj)} = \frac{\sum_{i \in P_k} x_{ki}}{\sum_{i \in p_k} w_{ki}^{**(hj)} x_{ki}} w_{ki}^{**(hj)} . \qquad (0.20)$$

This variance formula does not incorporate the first-phase finite population correction (fpc), $(1 - f_h)$ although the first-phase sampling is without-replacement and, therefore, overestimates the variance. However, if it is incorporated as given in the formula (0.6), the resulting variance estimator underestimates the variance because it corrects too much, ignoring the fact that subsampling of providers as well as nonresponses at the second phase occurred. The magnitude of the bias in either variance estimator largely depends on the overall sampling fraction, which is not substantial. To slightly understate the precision of estimates that were obtained from a sample with low response rate and sampling irregularities, we decided to use the formula (0.18) without incorporating the fpc. A correct variance estimator became available when weighting was nearly complete, so a study was conducted to make sure that the overestimation was not excessive. The study results and the correct variance estimator are presented in Appendix F.

Addressing Item Nonresponse

Item nonresponses for the service provider data were not imputed because item nonresponse was very low. We examined the data carefully before making a determination on the necessity of imputation. When the skipped cases were included in the calculation, item response rates were higher than 95 percent for all but two of the 658 variables (their response rates were 94.9 and 82.4 percent). In many cases, a

question was skipped because it was not applicable or irrelevant. No analysis was allowed for variables with less than 30 non-skipped respondents (i.e., the non-skipped rate should be at least 0.4 percent for data analysis). There were four variables with the non-skipped rate less than 0.4 percent. Since the skipped cases were not used in analyzing a variable, the response rate that was computed including the skipped cases as respondents was misleading. Therefore, the response rate was also examined excluding the skipped cases in the calculation. There were six variables for which the weighted or unweighted response rates (excluding the skipped cases) were below 90 percent. These variables are listed in Table 7. With the exception of SOTHMET3, SROCHESP, SOTHMET2, and SPOSNAME, the bias due to missing values would not be excessive, and the remaining six variables may also be used in analyses, but with caution.

Table 7. Variables with (1) Unweighted or Weighted Response Rates Below 90% After Removing Skipped Cases or (2) 30 or Fewer Non-Skipped Cases

Variable Name	Without Skipped Cases				Description of Variables
	Unweighted (%)	Weighted (%)	Difference (%)	No. of Non-skipped	
SOTHMET3	100.00	100.00	0.00	7	OTHER SPECIFY FROM SD8
SROCHESP	100.00	100.00	0.00	11	SD12-ROCHESTER METHOD
SOTHMET2	100.00	100.00	0.00	17	OTHER SPECIFY FROM SD8
SPOSNAME	96.66	98.64	-1.97	30	SA27-OTHER POSITION TITLE
SGREADPP	87.50	89.32	-1.82	208	SD7-TEACH READING IN GRADE YN
SPRESERV	71.25	59.29	11.95	480	SE27-RECEIVE ANY PRESERVICE PREPARATION
SLEPCLAR	81.71	85.02	-3.30	1,378	SD36-TEACH CONCEPT TO MAKE CLARIFICTN
SLEPLAGE	81.56	84.70	-3.13	1,378	SD36-USE NATIVE LANG TO TEACH ENG
SENCOURG	95.91	80.75	15.16	2,965	SD34-ENCOURAGE STUD INVOLVEMNT IN IEP
STURNOVR	80.20	87.26	-7.06	7,173	SB7-TEACHER HIRED TO REPLACE OTHERS

Imputation was done for almost all variables in the administrator survey. Various methods were used for imputation of missing values for the administrator sample depending on the type of variable (continuous or categorical), the availability of meaningful auxiliary information, the amount of missing data, the size of the donor pool, and the impact on the skip patterns that the imputed value had. The main methods of imputation are described in the following section. These methods were used with some variations when applied to a particular variable. Imputation actions taken for each variable are summarized in Appendix G.

Method 1 - Hot Deck Imputation

The entire data set was divided into a number of mutually exclusive imputation classes. For a missing value in an imputation class, a donor from the cases with no missing value in the same imputation class

was randomly selected, and the donor's value was imputed for the missing value. The imputation classes were defined by the design stratum cells. A donor was not used more than once to the extent possible. If there were not enough donors in an imputation class, the class was collapsed with a neighboring imputation class. The collapsing scheme differed from case to case. The general rule, however, was:

- collapse across the region within the same size group;
- if insufficient, collapse to a neighboring size group until enough donors are pooled;
- the type of agency in the sample (LEA, IEU, and State school) was not crossed.

This method was used mostly for categorical variables.

Method 2 - Nearest Neighbor (NN) Imputation

Instead of finding a donor randomly from an imputation class, as with the Hot Deck imputation method, a donor was selected using one or a few matching variables, often from the same imputation class as defined above. The nearest donor by the matching variables was used if it was unique; otherwise, a donor was randomly selected.

This method was used most often for continuous variables, where continuous auxiliary variables were available. In many cases, the auxiliary variables were not used directly but used to predict the variable being imputed for the recipient (the unit with the missing value) that would receive the donor value. This predicted value was used as the matching variable to compute the distance using the Euclidean measure.

A nearest donor was sought within the imputation class to the extent possible. Occasionally, the distance was too large, and the donor value was too different from the predicted value. In these cases, the recipient's predicted value was used to impute the missing value. When this method was applied to a dichotomous variable, logistic regression was used to define the matching variable.

Auxiliary variables used in this method were mainly QED district information on the total number of teachers employed, race/ethnicity of the teachers employed, and the roster information on service providers obtained during the recruitment process. The QED information on those variables was available for all LEAs and adequate for this use. However, it was not available for any State schools and only partially available for IEUs. Because data were largely missing for IEUs, other auxiliary variables were used. In sum, different sets of auxiliary variables were used for different types of agencies.

Sometimes survey variables were also used as the auxiliary variables. If the survey variables had missing values, they were imputed before using them as auxiliary variables. Therefore, an imputed value could be used for imputation of another variable. This situation is called composite imputation.

Regression modeling was performed using unweighted respondents for the variable under imputation within each agency type but disregarding the size groups in most cases. Initially, we attempted to use the sampling weights in the modeling, but this was discontinued because it did not make much difference in the imputation results.

Method 3 - Logistic Regression Imputation

When the variable with missing values was dichotomous (having a value of 1 or 2) and there were good auxiliary variables to predict the probability of 1, logistic regression was used. The logistic regression equation was estimated using the respondent data to predict the probability, and then, an experiment was performed using the predicted probability with a uniform random number generator between 0 and 1 to determine the value to be imputed. For example, if the predicted probability was 0.7 and the generated uniform random number was less than 0.7, then the value of 1 was imputed. Consequently, 70 percent of the time, the imputed value would be one, and 30 percent of the time, it would be 2. This method should be distinguished from the use of logistic regression to construct a matching variable for NN imputation.

Method 4 - Mean Imputation

When the variable was continuous and the number of missing values was very small, mean imputation was used. The mean of respondent values of the variable in an imputation class was used to impute any missing values in the class. This method was used only for a small number of cases because a distorted distribution can result when the number of missing values is high.

Method 5 - Logical Imputation

For a small number of cases, the missing value can be logically deduced from the reported values. For example, the six racial categories of special education teachers (question MA5) should add to 100 percent, and, if only one response category is missing, then the missing value should be 100 minus the sum of nonmissing response categories.

Method 6 - Imputation by Hand-picked Values

Several questions in the first section of the administrator survey asked about the correctness of roster information. The variables corresponding to these questions were imputed assuming that the roster information was correct. The number of missing values for these variables was very small (not more than three); however, their imputed values had far reaching impact on the skip patterns of many other variables. Thus, when the imputed values caused a change in the skip patterns of those cases, their imputed values were altered.

For questions in which respondents were given the option of responding "other" and specifying a nondesignated response, missing values were imputed with a value of 2 to avoid imputation of specified answers.

Method 7 - Mixed Method of Logistic Regression and Regression Imputation

When there were many cases with zero values for a continuous variable, the occurrence of zeros was modeled by logistic regression. Then, using the estimated logistic regression equation, it was determined whether a missing value was imputed by a zero or non-zero value. When a non-zero value was to be imputed, regression imputation was used.

Method 8 - Imputation for the Salary Schedule Variables

There were substantial missing values for information from agency salary schedules. Although we primarily planned to use salary schedules for 1999 and 2000 in the analysis, data for 1998 were also collected. Trend-adjusted values were used for missing 1999 or 2000 values whenever possible. However, this could take care of only a small number of cases. There were also many different missing patterns among those variables. Different imputation was carried out depending on the missing patterns--number of missing values and availability of 1998 data. When data for seven or more of the eight variables were missing, the 1999 base salary for a teacher with a BA (BABASE99) value was first imputed by ratio adjusting the 1998 value, if available, or if not available, by using random regression imputation. Then, a random donor was selected from the imputation class, and donor values were ratio-adjusted by the ratio of the recipient's BABASE99 value to the donor's BABASE99 value to impute other missing values. Similar donor imputation was performed when all 1999 values were present, but all 2000 values were missing or vice versa. The same approach was also used for two cases where only BABASE values were present, but all other variables were missing. When the number of missing values

was small, values were imputed using regression imputation based on nonmissing values as auxiliary variables.

A small number of odd cases that did not exactly follow the patterns described above for salary schedule imputation were forced to fit one of the main patterns. The method appropriate for that pattern was used.

Method 9 - Regression Imputation for Imputing the MNHLEP Variable

In order to impute the number of newly hired teachers qualified to work with limited English proficient (LEP) students (MNHLEP), a linear regression equation was estimated within each class defined by the agency type. Only two classes were needed as no missing value was found in the sample of IEUs. The auxiliary variable used in the regression was the number of full-time equivalent special education teachers employed in the agency who were qualified to work with LEP students (MFTELEP). The predicted values obtained from these equations were used to impute the missing values.

Results of Efforts to Reduce Nonresponse Bias

Although nonresponse weighting adjustment was carried out to minimize nonresponse bias, considering that the overall response rate for the provider data was very low (about 30 percent), SPeNSE estimates were compared with available external data as a means of quality assurance. Table 8 presents this comparison.

Analyzing the Data

Westat used a variety of analytic methods to describe and summarize the data collected through SPeNSE. This section of the document describes those methods. It includes analysis of single variables, including creating aggregate measures, and conducting bivariate and multivariate analyses. In conducting these analyses, Westat was guided by practices recommended by the National Center for Education Statistics (U.S. Department of Education, 1992).

To ensure that comparisons, conclusions, and inferences cited in descriptive reports were based on appropriate statistical procedures, Westat analyzed only those items for which there was an acceptable level of "response" in the survey. The following criteria were set for nonresponse. Results were not analyzed if the effective sample size was less than 20 on the administrator survey or less than 30 on the

service provider survey. For continuous variables, results were not analyzed if the CV was greater than 30 percent.

Table 8. Comparison of SPeNSE Estimates with External Data Sources

	SPeNSE		Other Source			Significant Difference[b]
	Estimate	95% CI	Estimate	95% CI	Source[a]	
Percent of non-White special education teachers	14.5%	11.4-18.2	14.0%	NA	1	No
Percent of special education teachers who are White	85.5%	81.8-88.6	86.0%	NA	3	No
Percent of special education teachers who are Black or African American	11.0%	8.2-14.5	10.0%	NA	3	No
Percent of special education teachers who are Hispanic	3.6%	2.4-5.1	2.0%	NA	3	Not Certain
Percent of special education teachers who are not White, Black or African American, or Hispanic	2.3%	1.3-3.0	2.0%	NA	3	No
Percent of full-time special education teachers not fully certified	8.0%	6.7-10.1	10.0%	population data	4	No
Percent of full-time general education teachers not fully certified	4.6%	3.8-7.5	7.0%	NA	5	No
Percent of special education teachers who work in regular preschool or K-12 schools	93.2%	92.0-94.4	95.0%	NA	3	Not Certain
Percent of school districts reporting they used state certification as a criterion for considering special education teaching applicants	100.0%	100-100	99.7%	NA	6	No
Average starting salary	$25,750	25184-26316	$25,888	25766-26010	2	No
Routine duties & paperwork interfere with teaching = agree or agree strongly	72.1%	68.8-75.2	71.2%	70.6-71.8	2	No
Percent of teachers who taught students with IEPs	85.8%	82.6-88.4	82.0%	81.3-82.7	2	Yes
Percent of teachers who taught LEP students	49.3%	44.8-53.9	41.2%	40.3-42.1	2	Yes
Percent of general education teachers with Master's degree	49.3%	45.7-55.4	45.0%	42.8-47.2	7	No
Percent of general education teachers whose professional development during the previous 12 months included working with parents	44.3%	41.0-47.7	46.0%	NA	7	No
Percent of general education teachers who taught students with disabilities who felt very confident in making educational decisions for them	29.1%	24.6-34.0	32.0%	NA	7	No

[a]See external source codes below.

[b]If an estimate from another source is contained within the 95 percent confidence interval for SPeNSE, the difference is not significant. If it is outside the confidence interval and its standard error is not available, it is not certain whether or not the difference is significant.

Table 8. Comparison of SPeNSE Estimates with External Data Sources (continued)

Source Code	Source
1	Kozleski, E., Mainzer, R.W., Deschler, D., Coleman, M.R., & Rodriguez-Walling, M. (2000). Bright futures for exceptional learners: An agenda to achieve quality conditions for teaching and learning. Arlington, Virginia [CEC statistic]
2	Schools and Staffing Survey (1999-2000)
3	Who's teaching our children with disabilities? (1995).NICHCY & NCPSE (http://www.kidsource.com/kidsource/content4/Spec_Ed/Spec_Ed.html). Based on 1990-91 data analyzed by Cook & Boe.
4	OSEP report to Congress - www.IDEAdata.org
5	Boe, E., & Barkanic, G. (2000, April). Influences on the supply and allocation of quality teachers: Analyses of the schools and staffing survey. Paper presented at the meeting of the American Education Research Association, New Orleans, LA.
6	Levine, R., & Christenson, B. (1998, February). Public school districts in the US: A statistical profile: 1987-88 to 1993-94. National Center for Education Statistics. Washington, DC. [SASS data]
7	Parsad, B., Lewis, L., & Farris, E. (2000). Teacher preparation and professional development. National Center for Education Statistics. Washington, DC.

Note that data sources 3 and 6 are fairly old, and the comparisons with these data sources should be taken with caution. Nevertheless, the comparison as a whole indicates that SPeNSE data are reasonably representative of the target population, and estimates obtained from the data are reasonably unbiased.

Analyzing Single Variables from SPeNSE Interviews

Westat used data from SPeNSE to paint a national portrait of the characteristics of personnel serving students with disabilities, policies affecting personnel, working conditions, preservice education, and continuing professional development. Descriptive statistics such as frequencies and distributions, means and proportions, medians, and standard deviations of these estimates were produced to summarize the SPeNSE data item by item and in combination.

Westat developed many derived variables for use in the descriptive and multivariate analyses. For example, Westat created a variable that categorized teachers who worked with children ages birth to 5, 6 to 11, 12 to 21, or combinations of those age ranges rather than analyzing responses related to the individual ages taught. Other areas in which Westat developed aggregate and derived variables include scales for school climate, skillfulness on job related tasks, use of best practices in managing behavior, promoting inclusion, teaching English language learners, teaching reading, and facilitating transition. Because these scales had varying numbers of items for different types of personnel, they were rescaled so possible scores ranged from 0 to 100. Appendix H includes the Chronbach's alpha for each of the scales developed as well as those that had inter-item correlations that were too low for constructing a scale. Items with a Cronbach alpha coefficient of .3 or higher were retained in each scale.

Descriptive analyses were presented for various subgroups, including all special education teachers, special education teachers who served primarily children ages 3-5, teachers serving students with visual or hearing impairments, teachers serving students with emotional disturbance, all other special education teachers, speech-language pathologists, special education paraprofessionals, and general education teachers.

Conducting Bivariate Analyses

Bivariate analyses were used to explore the relationships between two variables, for example, certification status by age or turnover rate by metropolitan status. Through cross-tabulations, chi-squares, analysis of variance, and other inferential statistics, analysts explored the relationships between important variables for all types of service providers. The bivariate analyses were used to describe variation in the characteristics of service providers across geographic areas, schools, and types of service providers. These bivariate analyses were chosen with input from stakeholders, the SPeNSE Advisory Group, and OSEP staff.

In the bivariate analyses, some variables were used to define subgroups. Subgroups included those defined based on:

- Type of service provider,
- More and less experienced teachers (less than 3 years and 3 or more years),
- Geographic region,
- District size,
- District metropolitan status,
- District wealth,
- Fully versus less than fully certified teachers,
- Teachers with and without a major in special education (BA or MA), and
- Grade/age level taught.

Credentials such as years of experience and certification were used as both dependent and independent variables. In analyses that used credentials as a dependent variable, Westat used service providers' demographic characteristics and school and district contextual measures as independent

variables. For example, we analyzed years of experience by race/ethnicity and certification status by district wealth.

While the SPeNSE administrator survey does not include many variables related to the supply of special education teachers, some aspects of supply were explored through information on shortages of teachers. Shortages were described through the proportion of positions left vacant or filled by a substitute teacher, the extent to which administrators reported an inadequate supply of teachers as a barrier to recruitment, and the number of person days of substitute teachers used in a typical week relative to the size of the agency. Since the study was not longitudinal, no information was available on the retention/attrition of individual service providers. Consequently, turnover issues could be explored only at the LEA, IEU, and State school levels, not at the service provider level. For service providers, intent to stay in the profession was used as a proxy for attrition.

Building Explanatory Models

Westat used several multivariate models to explain variation in the quality and turnover of personnel serving students with disabilities. These models explored relationships between a set of explanatory variables or independent variables, and the dependent or response variable. The primary dependent variables included intent to stay in teaching, manageability of teachers' workload, and a broader teacher quality measure. Guided by the results of the descriptive analyses, the bivariate analyses were extended to explore relationships between one dependent or response variable and a set of explanatory variables in the form of multiple regression, logistic, and loglinear regression models.

The quality of each service provider was not measured directly in SPeNSE, but many related variables that are associated with quality were measured. Since the survey variables are highly correlated with the latent quality variable, they can be used as proxy quality measures. Factor analysis was used to explore the construct of teacher quality. It attempted to identify common factors that may exist among the correlated variables. For major constructs associated with workforce quality, such as professional activities, classroom practices, and self-efficacy, multiple variables were used. Since many variables used in the factor analyses were ordinal, Westat used LISREL (Jöreskog & Sörbom, 1996), which allows for continuous and ordinal variables.

References

Boe, E.E., Bobbitt, S.A., Cook, L.H., Whitener, S.D., & Weber, A.L. (1997). Why didst thou go? Predictors of retention, transfer, and attrition of special and general education teachers from a national perspective. *The Journal of Special Education, 30*, 390-411.

Council for Exceptional Children. (1998). *CEC international standards for entry into professional practice* [On-line]. Available: http://www.cec.sped.org/ps/ps-entry.htm

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence*. Seattle: Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammond, L., & Ball, D.L. (1997). *Teaching for high standards: What policymakers need to know and be able to do* [On-line]. Available: http://www.negp.gov/Reports/highstds.htm

Ehrenberg, R.G., & Brewer, D.J. (1994). Do school and teacher characteristics matter?: Evidence from high school and beyond. *Economics of Education Review*, *13*, 1-17.

Ewing, N. (1995). Restructured teacher education for inclusiveness: A dream deferred for African American children. In B.A. Ford, F.E. Obiakor, & J. Patton (Eds.), *Effective education of African American exceptional learners* (pp. 189-208). Austin: Pro-Ed.

Ferguson, R. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation, 28,* 464-498.

Ferguson, P., & Womack, S.T. (1993). The impact of subject matter and education coursework on teaching performance. *Journal of Teacher Education*, *44*(1), 55-63.

Frechtling, J.A., Sharp, L., Carey, N., & Vaden-Kiernan, N. (1995). *Teacher enhancement programs: A perspective in the last four decades*. Washington, DC: National Science Foundation.

Garcia, S.B., & Malkin, D.H. (1993). Toward defining programs and services for culturally and linguistically diverse learners in special education. *Teaching Exceptional Children*, *26*(1), 52-58.

George, N.L., George, M.P., Gersten, R., & Grosenick, J. (1995). To leave or to stay: An exploratory study of teachers of students with emotional and behavioral disorders. *Remedial and Special Education, 16*, 227-236.

Goldhaber, D.D., & Brewer, D.J. (2000). Does teacher certification matter? *Education Evaluation and Policy Analysis*, *22*(2), 129-146.

Greenwald, R., Hedges, L.V., & Laine, R.D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*, 361-396.

Hanushek, E.A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro-data. *American Economic Review*, *61*, 280-288.

Ingersoll, R.M., Alsalam, N., Quinn, P., & Bobbitt, S. (1997). *Teacher professionalization and teacher commitment: A multilevel analysis* (NCES. No. 97-069). Washington, DC: National Center for Education Statistics.

Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: Users' reference guide*. Chicago: Scientific Software International, Inc.

Kim, J.K., Navarro, A., & Fuller, W. (2000). Variance estimation for 2000 census coverage estimates. In *Proceedings of the Survey Research Methods Section*, *American Statistical Association*. p 515-520.

Kott, P.S., & Stukel, D.M. (1997). Can the jackknife be used with a to-phase sample? *Survey Methodology*, 23, 81-89.

Khattree, R., & Naik, D.N. (2000). *Multivariate data reduction and discrimination with SAS software*. New York: Wiley.

Lee, H. (2001a). Technical note for weighting the SPeNSE provider data. Memo no. SPeNSE-STAT01-V2, Westat.

Lee, H. (2001b). SPeNSE: Overestimation of the variance. Memo no. SPeNSE-STAT08, Westat.

Mani, M.N.G. (1989). Analysis of the factors contributing to the retention of special education teachers in Colorado. (Doctoral Dissertation, University of Northern Colorado). *Dissertation Abstracts International, 50,* 07A.

Metzke, L.K. (1988). A study of the causes of teacher attrition in regular and special education in Wisconsin. (Doctoral Dissertation, Marquette University, WI). *Dissertation Abstracts International, 50,* 42A.

Morvant, M., Gersten, R., Gillman, J., Keating, T., & Blake, G. (1995). *Attrition/retention of urban special education teachers: Multi-faceted research and strategic action planning*. (ERIC Document Reproduction Service No. ED 338 154)

National Association of State Boards of Education. (1998). *The numbers game: Ensuring quantity and quality in the teaching work force*. Alexandria, VA: Author.

National Association of State Directors of Teacher Education and Certification. (2000). *Manual on certification and preparation of educational personnel in the United States (5th ed.)*. Dubuque, IA: Kendall/Hunt.

Singer, J. (1992). Are special educators' career paths special? Results from a 13-year longitudinal study. *Exceptional Children*, *59*, 262-279.

Singh, K., & Billingsley, B.S. (1996). Intent to stay in teaching. *Remedial and Special Education, 17*, 37-47.

SPSS (1993). *SPSS for Windows: CHAID Release 6.0*. Chicago: SPSS Inc.

Talbert, J.E., & McLaughlin, M.W. (1994). Teacher professionalism in local school contexts. *American Journal of Education*, *102*, 123-153.

U.S. Department of Education, National Center for Education Statistics. (1992). *Statistical standards* [On-line]. Available: http://nces.ed.gov/statprog/analysis.asp

Westat (2000). *WESVAR 4.0: User's Guide.* Rockville: Westat.