



February 25, 2000

NATIONAL LONGITUDINAL TRANSITION STUDY II (NLTS2)

SAMPLING PLAN

SRI Project 3421

Prepared by:
Renée Cameto
Mary Wagner
Lynn Newman
Jose Blackorby
Harold Javitz

NLTS2 is being designed under Task Order 18 under the Office of Special Education Programs' Initiative to Link Research and Practice to Improve Results for Individuals with Disabilities.

CONTENTS

Sample Parameters	1
General Sampling Approach	2
Student Sample Size	2
The LEA Sample	5
LEA Sample Size	5
Defining the Universe of LEAs	6
Creating the Sampling Frame	7
Stratification	8
The Stratified Universe	10
LEA Sample Characteristics	12
Weighting	15
Student Sample Selection Procedures	16
Minimizing Sample Attrition	17
Reference	18

TABLES

Table 1	Approximate Number of Students Receiving Special Education Ages 13 to 16 in U.S. Public Schools, by Disability Category	4
Table 2	Expected Sample Size by Year and Disability Category	5
Table 3	Distribution of States by Region	8
Table 4	Distribution of LEAs and Student Population by Region	9
Table 5	Distribution of LEAs and Student Population by LEA Size	10
Table 6	Distribution of LEAs and Student Population by District Wealth	10
Table 7	Number of LEAs in the Universe/Sample, by Stratum	11
Table 8	Number of Students in the Universe/Sample of LEAs, by Stratum	12
Table 9	Weighted and Unweighted Distribution of Sampled LEAs and Student Population, by Region	13
Table 10	Weighted and Unweighted Distribution of Sampled LEAs and Student Population, by LEA Size	13
Table 11	Weighted and Unweighted Distribution of Sampled LEAs and Student Population, by District Wealth.....	14
Table 12	Weighted Distribution of Sampled LEAs and Universe, by Metropolitan Status	15
Table 13	Weighted Distribution of Sampled LEAs and Universe, by Proportion of Minority Students	15
Table 14	Estimated Student Sampling Fractions, by LEA Size Stratum	16
Table 15	Expected Student Sampling Weights, by Size Strata	17

NLTS2 SAMPLING PLAN

Sample Parameters

NLTS2 must meet the information needs of a wide variety of audiences using a variety of data collection and analytic approaches. The NLTS2 sample must meet the following requirements in order to serve its multiple purposes:¹

- **Focus on students.** NLTS2 data must produce accurate estimates about the characteristics, programs, and outcomes of students receiving special education. However, no list of all students receiving special education exists from which to draw the NLTS2 sample. Thus, a sample of LEAs must be drawn, from whose rosters students can be selected. However, the sample of LEAs is only a vehicle to obtaining a sample of students; it is too small to make highly precise national estimates about LEA practices (OSEP has commissioned a separate study of state and local implementation of IDEA97 to meet this latter purpose).
- **Generalize to each disability category.** The NLTS2 sample must enable reasonably precise estimates for the various disability categories of the special education student population ages 13 through 16 who are in at least 7th grade.
- **Longitudinal.** NLTS2 data will be collected repeatedly over a 9-year period. The initial sample must be large enough to support estimates of reasonable precision in the ninth year of data collection (assuming that 8% of students who are in the sample each year will be lost the following year because of mobility).²
- **Multiple data sources.** Multiple data sources will be needed to obtain the breadth of information specified in the NLTS2 conceptual framework (see SRI, 2000). Many analyses will employ information from more than one source. Some students will not have information from a source, reducing the sample for analyses using that data source. Even more will be missing information when several sources are combined. The sample must be large enough to accommodate missing information from multiple data sources.
- **Multiple analytic purposes.** The richness of the NLTS2 database will support a variety of analyses that have implications for the sample design. For example, subgroup analyses will examine experiences and outcomes of students receiving special education who are differentiated by particular characteristics (other than age and disability category), such as gender, ethnicity, or functional abilities. The NLTS2 sample must be large enough to support such subgroup analyses.

¹ Throughout this discussion of the NLTS2 sampling approach, we have made a variety of assumptions regarding statistical precision, attrition, response rates, etc. We have based many of these assumptions on the National Longitudinal Transition Study. Although we believe these are reasonable assumptions, other contractors could use different assumptions that would result in different sample size estimates.

² The assumption of 8% attrition reflects experience with the National Longitudinal Transition Study, in which aggressive tracking efforts kept sample attrition to about 6% per year. Changing demographics suggest that a higher attrition rate may be experienced in NLTS2.

- Comparable to NLTS. The sample must permit comparisons with the original NLTS, conducted by SRI from 1985 through 1993 in order to determine changes in the experiences and achievements of students in transition over the past decade or more.

In the remainder of this section, the approach to meeting these sample requirements is presented.

General Sampling Approach

NLTS2 will employ a two-stage process to generate the needed sample of students receiving special education between the ages of 13 and 16 who are in at least 7th grade. NLTS2 will draw a random sample of students receiving special education from a nationally representative sample of LEAs and a sample of state-supported special schools. Accordingly, the LEA is the primary sampling unit and the student is the secondary or final unit.

The NLTS2 sample will be generated by randomly selecting students receiving special education from rosters of LEAs and state-supported schools that serve students of the appropriate ages in special education. The universe of eligible LEAs and special schools will be stratified by key factors to enhance representativeness; these factors are geographic region, district enrollment, and district/community wealth. Taking into account the length of the data collection period and assumptions regarding attrition from the sample, analyses of statistical power requirements suggest that an initial sample of approximately 11,500 students will yield a sample of sufficient size and representativeness to meet the analytic needs of NLTS2 in its final wave of data collection. This sample will be selected so as to generate 1,250 in each disability category, with the exception of the three least populous categories—1,012 students who are autistic, 559 with traumatic brain injuries, and 122 who are deaf-blind.

The following sections describe the process through which the student sample size was determined and then outline the selection procedures for the LEA and student samples.

Student Sample Size

The size of the NLTS2 student sample is a function of the duration of the study, desired levels of precision, and assumptions regarding attrition and response rates. The following assumptions have been used in determining the size of the student sample:

- Location information (parent name, address, telephone number) will be provided by LEAs for 90% of sampled students.³ Therefore, for each 1,000 students sampled in year 1, location information will be available for 900 students.
- In each year of the study after the first year, 92% of the students from the preceding year will be retained. Thus, for each 1,000 students sampled in year 1, 900 will have

³ Experience from NLTS has demonstrated that some LEAs will not reveal location information for students. In such cases, the contractor is likely to need to provide letters of invitation for parents to the district, whose staff will then mail them to parents. Only parents who elect to return the consent form, thus identifying themselves, can be included in data collection.

location information and the sample can be expected to retain 762 in year 3, 645 in year 5, 546 in year 7, and 462 in year 9.

- For each 1,000 students sampled in year 1, parent/guardian or student interviews will be completed for 70%⁴ of students retained in the sample, or 630 students in year 1 (i.e. 70% of the 900 students whom we can track), 533 in year 3, 452 in year 5, 382 in year 7, and 323 in year 9.
- Direct assessments will be completed for 75% of the youth retained in the sample each year. Therefore, for each 1,000 students age 15, and 16 who are sampled in year 1, assessments will be completed for 621 students in year 2 (i.e., $1000 \times .90 \times 0.92 \times 75\%$). For each 1,000 students age 13 or 14 who are sampled in year 1, assessments will be completed for 526 students in year 4.
- The number of students who have both a parent/guardian interview in year 1 and a direct assessment in year 2 will be 75% x 92% of the number of students who have a parent/guardian interview in year 1 (i.e., 435 students for each 1,000 sampled in year 1).

The NLTS2 sample design emphasizes the need to estimate proportions and ratios (for example, the percentage of students receiving special education who enroll in post-secondary education) instead of estimating the actual numbers of students receiving special education having specified characteristics (for example, 134,400 students who enroll in liberal arts colleges). However, relatively precise national estimates of the proportions or ratios of students receiving special education, whether analyzed as one group or considered separately by disability category, will be needed to adequately answer research questions of interest to the broad range of likely audiences for the study.

After consideration of various options, a target standard error of 3.6% in year 9 parent/youth interviews for the most populous disability categories was selected for the NLTS2. On the basis of the sample design and the experience of NLTS (wherein a sampling efficiency of approximately 50% was achieved), we expect the design to achieve this target for the categories of learning disabilities, speech impairments, serious emotional disturbances, mental retardation, hearing impairments, and other health impairments. The target is almost achievable for the categories of visual impairments (3.8%) and multiple impairments (3.8%). Precision targets for the remaining disability categories are lower because of their relatively low prevalence—orthopedic impairments (4.1%), autism (6.1%), traumatic brain injury (8.2%), and deaf-blindness (10.1%).

Expansion of the sample size to achieve a target standard error of 3.6% for all disability categories would be prohibitively expensive, particularly given the central importance and considerable expense of a direct assessment of students. A sample of that size would be a sizable proportion of all the students, given the low-incidence of some disability categories (see Table 1). Students receiving special education account for approximately 10.6% of all students in American schools, the number of students ages 13 through 16 in each disability category ranges from a high of approximately 1,167,000 for students with learning disabilities (approximately 5.8% of the total student population) to a low of approximately 467 for deaf-blind students (far

⁴ This percentage is based on experience with the NLTS, which obtained parent interviews from 69.6% of households for whom location information was available.

less than 1% of the total student population). For example, to select the approximately 1,250 students necessary to reach a precision level of 3.6% for those with traumatic brain injury (TBI) would require selecting an extremely large number of LEAs (i.e., enough LEAs to encompass approximately 29% of the total student population).

Table 1 APPROXIMATE NUMBER OF STUDENTS RECEIVING SPECIAL EDUCATION AGES 13 TO 16 IN U.S. PUBLIC SCHOOLS, BY DISABILITY CATEGORY		
	Number of Students	Approximate Percentage of Student Population Ages 13 to 16
Learning disabilities	1,167,204	5.84
Speech impairments	74,231	0.37
Mental retardation	241,925	1.21
Serious emotional disturbances	230,081	1.15
Other health impairments	56,893	0.28
Multiple disabilities	31,927	0.16
Hearing impairments	25,756	0.13
Visual impairments	9,950	0.05
Orthopedic impairments	21,277	0.11
Autism	7,823	0.04
Traumatic brain injury	4,322	0.02
Deaf-blindness	467	0.00
Total	1,871,856	9.37

Table 2 shows the number of youth who are expected to be retained in the study for each year and for whom data are expected to be collected, based on a starting sample of 1,250 students in each category, with the exception of 1,012 students with autism, 559 students with traumatic brain injury and 122 with deaf-blindness. As a result of the desire to track post-secondary educational achievement, the sampling rates will be somewhat higher for students who are initially 16 years old than for those who are 13 to 15 years old. Whenever possible, sampling rates will be 50% higher for 16 year olds than for 13 to 15 year olds. This oversampling rate can be achieved for the categories of learning disabilities (LD), serious emotional disturbances (SED), and mental retardation (MR). Oversampling rates for older students for the other disability categories will be smaller because in some LEAs, it is necessary to select nearly all of the students in some disability categories at all ages. Table 2 shows the expected sample size for the three disability categories that have fewer than 1,250 students, and for the LD category.⁵ The results for the SED and MR categories are expected to be the same as for the LD category; the

⁵ Although the sampling fraction for older students is increased, the actual number of 16 and 17 year olds is still smaller than the number of younger students because the older age group has only two cohorts of youth and because there is a decrease in the number of youth in each cohort as they age.

results for the remaining disability categories will be the same as the LD category, except that the number of students in each age group will be adjusted proportionately.

Table 2				
EXPECTED SAMPLE SIZE, BY YEAR AND DISABILITY CATEGORY				
	Autism	TBI	Deaf-blindness	Learning Disabilities
Number of students				
Sampled, age 13 to 15	688	332	73	710
Sampled, age 16	324	227	49	540
Total Sampled	1,012	559	122	1,250
With location information	911	503	110	1,125
Year 1				
Year 3	771	426	93	952
Year 5	653	361	79	806
Year 7	552	305	67	682
Year 9	468	258	56	577
Number of parent/guardian or student interviews				
Year 1	634	350	77	783
Year 3	537	297	65	663
Year 5	454	251	55	561
Year 7	385	212	46	475
Year 9	325	180	39	402
Age 24 or 25 at last interview	104	73	16	174
Number of direct assessments				
Year 2 (initially ages 16)	107	76	16	158
Year 4 (initially ages 13 to 15)	226	115	25	247

The LEA Sample

The first step in developing a sample that leads to national estimates about students receiving special education is to select an adequate, representative sample of LEAs. Below we discuss issues related to the LEA sample including size, stratification, and fit.

LEA Sample Size

There are several factors to consider in determining the number of LEAs for the sample. First, it is necessary to establish the number of LEAs that are required to generate the needed student sample. On the basis of an analysis of LEAs' estimated enrollment across district size, and estimated sampling fractions for each disability category, 497 LEAs (and as many state-sponsored special schools as will participate) will be sufficient to generate the student sample. Second, the

rate of LEA refusal to participate should be considered so that the required number of LEAs agree to participate within the limited recruitment period and budget. Previous experience with NLTS suggests that LEAs typically declined to participate because of concerns related to confidentiality of student records. Although considerable time and effort was expended in recruiting LEAs for NLTS, approximately 55% of the LEAs invited to participate either declined, did not respond, or introduced procedures that unacceptably lengthened the recruitment process. In the Special Education Elementary Longitudinal Study (SEELS), both the amount of time and the funds available to recruit LEAs were less than were available in NLTS, and the recruitment rates were lower. In SEELS, approximately 48% of very large LEAs, 27% of large LEAs, and approximately 23% of medium and small LEAs agreed to participate.

In NLTS2 we have assumed the same LEA participation rate as in SEELS. Recruitment efforts will focus on the very large LEAs, which are relatively few in number and from which a relatively large proportion of sample students will be selected. Smaller LEAs will receive less intensive recruitment effort than in NLTS because there are many of them, yielding a large number of potential replacements for refusing districts. Although this strategy is likely to be most efficient in selecting the LEA sample quickly, there is a risk that smaller LEAs who refuse to participate differ systematically from other LEAs in terms of the types or effectiveness of programs that they offer to students. Thus, detailed tracking will be necessary to identify potential patterns that emerge with regard to LEA refusal/nonresponse. The procedural outcome of concentrating the recruitment effort on larger LEAs and being more willing to replace smaller LEAs is that a sample of 2,205 LEAs is expected to be required to generate 497 participating LEAs.

Defining the Universe of LEAs

The initial task in selecting the NLTS2 sample is to define which districts should be included in and excluded from the universe of LEAs from which the sample will be selected. To meet its purposes, the NLTS2 sample includes only LEAs that have teachers, students, administrators and operating schools—that is, “operating LEAs.” The NLTS2 sample excludes the following categories of local and state educational “districts” that appear on standard listings of educational institutions:

- Nonoperating LEAs, which do not administer any schools.⁶
- Vocational-technical districts (except those that operate as regular LEAs). These districts often are not comparable to LEAs in enrollment, operating hours, or administrative structure, making their inclusion problematic.
- Supervisory unions, area educational agencies, interim districts, boards of county education services, or other superordinate units. These organizations occur most frequently in rural areas where the individual district-level enrollments are quite small. There is evidence that the operation of superordinate units varies from state to state. For example, in many states, the local districts are fiscally responsible for students attending such service units, and they are therefore listed on local district

⁶ According to NCES, an operating system is a self-contained local public school system having its own decision-making board of control, operating a school or schools providing general elementary/middle school/secondary education. A nonoperating system is a self-contained local public school system having its own decision-making board of control, which does not operate schools but pays tuition to other operating systems for the education of the children living within its boundaries.

rosters. Such students would be included in NLTS2. In other states, such units may have fiscal responsibility for their students. However, we believe that the exclusion of such units is justified because (1) their inclusion introduces the risk of double counting in states where students are found on local rosters and on superordinate rosters, and (2) too few students are in such units for them to serve as an analytic category.

- Public agencies, such as state education agencies (with the exception of the Department of Education in Hawaii, which is an LEA); Bureau of Indian Affairs (BIA) schools; achievement centers and regional resource centers; private agencies, such as homes for delinquent students; and Texas Independent State School Districts, which primarily are correctional facilities and homes for delinquent students. Included, however, are the “accommodation” school districts in Arizona, which are regular operating LEAs with nontraditional boundaries (e.g., around federal dams and military installations).
- LEAs from Puerto Rico, Guam, and other territories, to reduce the cost and complexity of future data collection.
- LEAs that do not serve students in grades 7 through 12, which are most likely to encompass the age range of NLTS2 students (i.e., to be part of the sampling universe, an LEA must offer instruction in at least one of these grades).
- LEAs (most with very small enrollments) for which the stratifying variable of district wealth cannot be obtained (see “Stratification” section for a discussion of the stratification variables).
- LEAs with 10 or fewer students in grades 7 through 12. Such schools would have an estimated enrollment of less than one student in special education in the target age range.

Creating the Sampling Frame

To create a sampling frame or master list of LEAs, the school and agency universe maintained by Quality Education Data (QED) was used. As a commercial source, it maintains fairly accurate data, including addresses of special education coordinators in each district, for its clients. We used the most recent version of the QED database, released December 1999, which contains data from the 1998-99 school year, as updated during the fall of 1999. The following procedures were used to create a master list of LEAs that were eligible for the NLTS2 sample:

- Obvious errors were corrected, such as blank or duplicate records, no names, spelling errors, invalid codes, and extreme outliers.
- All nonoperating LEAs, supervisory unions, vocational-technical districts, and relevant public agencies were eliminated (see previous discussion), as were all districts that did not serve any grade in the grade 7 through grade 12 range.

These procedures resulted in a master list of 12,435 LEAs that are expected to have at least one student receiving special education in the appropriate age range. These comprise the NLTS2 LEA sampling frame.

Stratification

The NLTS2 LEA sample is stratified for four principal reasons: (1) to increase the precision of estimates by eliminating between-strata variance, (2) to ensure that low-frequency types of LEAs (e.g., large urban districts) are adequately represented in the sample, (3) to improve comparisons with the findings of other research, and (4) to make NLTS2 responsive to concerns voiced in policy debate (e.g., differential effects of federal policies in particular regions, LEAs of different sizes). The first of these reasons is especially important because of the great diversity in the universe of LEAs. Three stratifying variables are used—geographic region, district size (student enrollment), and a measure of district/community wealth. They were selected on the basis of conceptual soundness and the likelihood of providing a gain in precision over simple random sampling. These variables and their sources are described below.

Region. This variable captures essential political differences, as well as subtle differences in the organization of schools, the economic conditions under which they operate, and the character of public concerns. Regions differ, for example, in the changes in school enrollment over time. They also differ in terms of economic health, which is linked to resources the region can target to education and other needed services. For NLTS2, the regional classification variable selected is used by the Department of Commerce, the Bureau of Economic Analysis, and the National Assessment of Educational Progress (see Table 3).

Table 3		
DISTRIBUTION OF STATES BY REGION		
Northeast (N = 12)		
Connecticut	Maryland	New York
Delaware	Massachusetts	Pennsylvania
District of Columbia	New Hampshire	Rhode Island
Maine	New Jersey	Vermont
Southeast (N = 12)		
Alabama	Kentucky	South Carolina
Arkansas	Louisiana	Tennessee
Florida	Mississippi	Virginia
Georgia	North Carolina	West Virginia
Central (N = 12)		
Illinois	Michigan	North Dakota
Indiana	Minnesota	Ohio
Iowa	Missouri	South Dakota
Kansas	Nebraska	Wisconsin
West/Southwest (N = 15)		
Alaska	Idaho	Oregon
Arizona	Montana	Texas
California	Nevada	Washington
Colorado	New Mexico	Wyoming
Hawaii	Oklahoma	Utah

By assigning each LEA to a region based on its state, we obtain the allocation to region of LEAs and proportion of total estimated middle and high school student population in grades 7 through 12 that is indicated in Table 4.

Table 4				
DISTRIBUTION OF LEAS AND STUDENT POPULATION BY REGION				
Region	Number of LEAs	Percent of LEAs	Number of Students	Percent of Students
Northeast	2,525	20.3	3,800,844	19.0
Southeast	1,590	12.8	4,921,633	24.6
Central	4,656	37.4	4,706,463	23.6
West/Southwest	3,664	29.5	6,551,205	32.8
TOTAL	12,435	100.0	19,980,145	100.0

District size (student enrollment). LEAs vary considerably by size, the most useful available measure of which is pupil enrollment. A host of organizational and contextual variables are associated with size that exert considerable potential influence over the operations and effects of special education and related programs. These include the extent of district administrative/supportive capacity, the degree of specialization in administrative structure, the nature of citizen and interest group activity in education, and the characteristics of relationships with state and federal governance systems.

In addition, total enrollment (and the previously described estimated middle/high school enrollment) serves as an initial proxy for the number of students receiving special education served by a district. The QED database provides enrollment data from which LEAs were sorted into four categories serving approximately equal numbers of students:

- **Very large** (estimated enrollment greater than 14,931 in grades 7 through 12). These are either districts in large urban centers or large county systems, which are typically organizationally complex and likely to be divided into subdistricts.
- **Large** (estimated enrollment from 4,661 to 14,931 in grades 7 through 12). These are districts set in small to medium-sized cities or large county systems. They are also organizationally complex, but these systems tend to be centralized.
- **Medium** (estimated enrollment from 1,568 to 4,660 in grades 7 through 12). These typically are suburban districts, large rural towns, and small county systems.
- **Small** (estimated enrollment from 11 to 1,567 in grades 7 through 12). The majority of districts in the country fall into this group. Most are small rural districts, which most likely receive little money for special education programs; the range of activities that these funds can be used for is likely to be extremely narrow.

The distribution of districts among these strata and proportion of students accounted for by each stratum are displayed in Table 5.

Table 5				
DISTRIBUTION OF LEAS AND STUDENT POPULATION BY LEA SIZE				
Enrollment Size Category	Number of LEAs	Percent of LEAs	Number of Students	Percent of Students
Very large (>14,931)	141	1.1	4,984,021	24.9
Large (4,661 - 14,931)	646	5.2	5,009,778	25.1
Medium (1,568 - 4,660)	1,966	15.8	4,992,149	25.0
Small (11 - 1,567)	9,682	77.9	4,994,197	25.0
TOTAL	12,435	100.0	19,980,145	100.0

District/community wealth. LEAs differ greatly in the resources they have available and in the demands placed on those resources by low-income students whose needs put them at risk for a variety of problems, including school failure. Policies and programs may differ in LEAs that face these differential demands of disadvantaged students. Also, prior research has demonstrated that high-poverty districts have a high proportion of students receiving special education. As a measure of district wealth, the Orshansky index (the proportion of the student population living below the federal definition of poverty) is a well-accepted measure. The distribution of Orshansky index scores was organized into four categories of district/community wealth, each containing approximately 25% of the student population in grades 7 through 12:

- High (0% to 13% Orshansky)
- Medium (14% to 24% Orshansky).
- Low (25% to 43% Orshansky).
- Very low (over 43% Orshansky).

The distribution of districts among strata and proportion of students accounted for by each stratum are displayed in Table 6.

Table 6				
DISTRIBUTION OF LEAS AND STUDENT POPULATION BY DISTRICT WEALTH				
District Wealth (Orshansky Index)	Number of LEAs	Percent of LEAs	Number of Students	Percent of Students
High (0% – 13%)	3,612	29.0	4,723,187	23.6
Medium (14% – 24%)	3,070	24.7	5,256,479	26.3
Low (25% – 43%)	3,507	28.2	5,072,592	25.4
Very low (> 43%)	2,246	18.1	4,927,887	24.7
TOTAL	12,435	100.0	19,980,145	100.0

The Stratified Universe

The three variables generate a 64-strata grid into which the entire universe can be fit. Table 7 shows the strata and the number of LEAs in each stratum. Table 8 shows the number of students in all LEAs in each stratum. The next stage in the NLTS2 sampling process was to select the

appropriate LEAs from each stratum to yield a total sample of 2,205 LEAs.⁷ LEAs were selected from strata so as to maximize the sampling efficiency and thereby to maximize the effective sample sizes. To reduce respondent burden, the LEA sample was selected minimizing the overlap between the NLTS2 sample and the SLIIDEA sample. Table 8 indicates the number of students in LEAs selected for the sample in each stratum.

Table 7					
NUMBER OF LEAS IN THE UNIVERSE/SAMPLE, BY STRATUM					
District Size/Region	District Wealth (Orshansky Index)				Total
	High (0% – 13%)	Med (14% – 24%)	Low (25% – 43%)	Very Low (> 43%)	
Very large	15/7	37/18	34/16	55/28	141/69
Northeast	2/2	6/4	1/1	5/5	14/12
Southeast	4/2	11/5	21/11	15/7	51/25
Central	1/1	1/1	0/0	15/7	10/5
West/Southwest	8/2	19/8	12/4	27/13	66/27
Large	174/131	162/127	177/139	133/114	646/511
Northeast	26/17	21/14	9/6	14/12	70/49
Southeast	19/16	44/33	71/57	37/31	171/137
Central	53/37	38/27	29/22	14/14	134/100
West/Southwest	76/61	59/53	68/54	68/57	271/225
Medium	637/218	518/178	487/168	324/111	1,966/675
Northeast	268/87	133/44	64/23	25/9	490/163
Southeast	26/10	82/30	170/61	180/60	458/161
Central	255/89	172/57	95/30	27/10	549/186
West/Southwest	88/32	131/47	158/54	92/32	469/165
Small	2,786/294	2,353/257	2,809/254	1,734/145	9,682/950
Northeast	971/123	535/69	374/39	71/7	1,951/238
Southeast	41/6	93/13	339/49	437/61	910/129
Central	1,325/132	1,228/130	1,112/91	298/16	3,963/369
West/Southwest	449/33	497/45	984/75	928/61	2,858/214
TOTAL	3,612/650	3,070/580	3,507/577	2,246/398	12,435/2,205
Northeast	1,267/229	695/131	448/69	115/33	2,525/462
Southeast	90/34	230/81	601/178	669/159	1,590/452
Central	1,634/259	1,439/215	1,236/143	347/43	4,656/660
West/Southwest	621/128	706/153	1,222/187	1,115/163	3,664/631

⁷ A SAS computer program was written to select eligible LEAs from the QED database, sort them into the appropriate strata, and then randomly select LEAs from within each stratum. In practice, each LEA within a stratum was assigned a number and the program was instructed to draw a random number and match it to the appropriate LEA. This process was repeated until the full sample of LEAs was complete.

Table 8					
NUMBER OF STUDENTS IN THE UNIVERSE/SAMPLE OF LEAS, BY STRATUM (THOUSANDS)					
District Size/Region	District Wealth (Orshansky Index)				Total
	High (0% – 13%)	Med (14% – 24%)	Low (25% – 43%)	Very Low (> 43%)	
Very large	298/136	1,348/783	1,131/456	2,208/1,413	4,984/2,787
Northeast	33/33	271/220	55/55	521/521	880/829
Southeast	81/44	362/161	769/293	506/305	1,717/803
Central	18/18	158/158	0/0	506/305	399/239
West/Southwest	167/41	557/244	307/108	957/522	1,988/916
Large	1,274/946	1,243/948	1,364/1014	1,128/945	5,010/3,852
Northeast	163/106	135/82	57/38	118/95	472/320
Southeast	155/131	320/223	558/407	301/244	1,335/1,005
Central	359/237	266/189	219/156	148/148	992/730
West/Southwest	597/473	522/454	530/413	562/458	2,210/1,798
Medium	1,610/543	1,314/432	1,245/423	824/272	4,992/1,671
Northeast	641/210	324/97	169/59	67/22	1,201/388
Southeast	74/29	219/78	453/161	447/147	1,193/415
Central	656/221	423/134	224/70	72/24	1,375/448
West/Southwest	239/84	347/123	398/134	238/79	1,223/419
Small	1,542/182	1,352/145	1,333/126	768/79	4,994/532
Northeast	646/92	361/40	203/21	37/5	1,247/159
Southeast	29/4	68/9	257/36	323/48	677/98
Central	693/67	686/75	476/37	86/5	1,941/185
West/Southwest	174/19	237/20	396/31	323/22	1,130/91
TOTAL	4,723/1,806	5,256/2,308	5,073/2,019	4,928/2,709	19,980/8,842
Northeast	1,483/441	1,091/439	484/172	743/643	3,801/1,695
Southeast	339/207	969/472	2,037/896	1,576/745	4,922/2,321
Central	1,726/543	1,533/555	919/264	528/241	4,706/1,602
West/Southwest	1,176/616	1,663/841	1,632/686	2,080/1,081	6,551/3,224

LEA Sample Characteristics

Our first step in assessing the effectiveness of the sampling process was to evaluate the degree to which the selected LEA sample was comparable to the universe from which it was drawn on variables used in the sampling process. Tables 9, 10, and 11 depict the characteristics of the LEA sample, in weighted and unweighted form, on the sampling variables of region, LEA size, and LEA wealth. Taken together, the tables illustrate that the weighted LEA sample closely resembles the LEA universe with respect to those variables.

Table 9				
WEIGHTED AND UNWEIGHTED DISTRIBUTION OF SAMPLED LEAS AND STUDENT POPULATION, BY REGION				
	Number of LEAs	Percent of LEAs	Number of Students	Percent of Students
Region, Weighted				
Northeast	2,525	20.3	3,850,116	19.3
Southeast	1,590	12.8	4,774,048	24.0
Central	4,656	37.4	4,559,225	22.9
West/Southwest	3,664	29.5	6,716,231	33.8
TOTAL	12,435	100.0	19,899,621	100.0
Region, Unweighted				
Northeast	462	21.0	1,695,214	19.2
Southeast	452	20.5	2,320,719	26.2
Central	660	29.9	1,602,320	18.1
West/Southwest	631	28.6	3,224,019	36.5
TOTAL	2,205	100.0	8,842,272	100.0

*Compare with Table 4.

Table 10				
WEIGHTED AND UNWEIGHTED DISTRIBUTION OF SAMPLED LEAS AND STUDENT POPULATION, BY LEA SIZE				
	Number of LEAs	Percent of LEAs	Number of Students	Percent of Students
Enrollment Size Category, Weighted				
Very large (>14,931)	141	1.1	5,095,164	25.6
Large (4,661 - 14,931)	646	5.2	4,822,142	24.2
Medium (1,568 - 4,660)	1,966	15.8	4,857,109	24.4
Small (11 - 1,567)	9,682	77.9	5,125,204	25.8
TOTAL	12,435	100.0	19,899,621	100.0
Enrollment Size Category, Unweighted				
Very large (>14,931)	69	3.1	2,787,028	31.5
Large (4,661 - 14,931)	511	23.2	3,852,256	43.6
Medium (1,568 - 4,660)	675	30.6	1,670,676	18.9
Small (11 - 1,567)	950	43.1	532,312	6.0
TOTAL	2,205	100.0	8,842,272	100.0

*Compare with Table 5.

Table 11				
WEIGHTED AND UNWEIGHTED DISTRIBUTION OF SAMPLED LEAS AND STUDENT POPULATION, BY DISTRICT WEALTH (ORSHANSKY INDEX)				
	Number of LEAs	Percent of LEAs	Number of Students	Percent of Students
District Wealth, Weighted				
High (0% – 13%)	3,612	29.0	4,815,560	24.2
Medium (14% – 24%)	3,070	24.7	5,180,734	26.0
Low (25% – 43%)	3,507	28.2	4,769,822	24.0
Very low (> 43%)	2,246	18.1	5,133,504	25.8
TOTAL	12,435	100.0	19,899,621	100.0
District Wealth, Unweighted				
High (0% – 13%)	650	29.5	1,806,494	20.4
Medium (14% – 24%)	580	26.3	2,308,083	26.1
Low (25% – 43%)	577	26.2	2,018,565	22.8
Very low (> 43%)	398	18.0	2,709,130	30.6
TOTAL	2,205	100.0	8,842,272	100.0

*Compare with Table 6.

In addition to ensuring that the LEA sample matched the universe of LEAs on variables used in sampling, it was important to ascertain whether this stratified random sampling approach resulted in skewed distributions on relevant variables not included in the stratification scheme. Two variables from the QED database were chosen to compare the “fit” between the first-stage sample and the population: the district’s metropolitan status (Table 12), and the district’s proportion of minority students (Table 13). If comparisons between the universe of LEAs and the sample revealed a poor fit, either the sample would be reweighted or a new sample would need to be selected.

Tables 12 and 13 reveal that the fit between the weighted LEA sample and the LEA universe is quite good with respect to metropolitan status and the percent of minority students.

Table 12				
WEIGHTED DISTRIBUTION OF SAMPLED LEAS AND UNIVERSE, BY METROPOLITAN STATUS				
District Type	Number in Universe	Percent of Universe	Weighted Number in Sample	Percent of Weighted Sample
Unclassified	706,899	3.5	797,231	4.0
Large central city	2,436,343	12.2	2,580,814	13.0
Midsized central city	3,490,044	17.5	3,142,262	15.8
Urban fringe of large city	3,335,515	16.7	3,105,687	15.6
Urban fringe of midsized city	2,110,189	10.6	2,153,164	10.8
Large town	643,008	3.2	641,000	3.2
Small town	4,673,885	23.4	4,965,734	25.0
Rural	2,584,262	12.9	2,513,726	12.6
TOTAL	19,980,145	100.0	19,899,618	100.0

Table 13				
WEIGHTED DISTRIBUTION OF SAMPLED LEAS AND UNIVERSE, BY PROPORTION MINORITY STUDENTS				
Minority Student Population	Number in Universe	Percent of Universe	Weighted Number in Sample	Percent of Weighted Sample
Less than 5%	3,540,857	17.7	3,631,412	18.2
5% – 10%	2,281,071	11.4	2,286,414	11.5
10% – 20%	2,762,306	13.8	2,771,804	13.9
20% – 50%	5,367,052	26.9	5,128,687	25.8
50% – 100%	6,028,859	30.2	6,081,302	30.6
TOTAL	19,980,145	100.0	19,899,621	100.0

Weighting

Because LEAs have an unequal probability of being selected into the sample, depending on the stratum within which they fall, LEAs need to be weighted by the inverse of the stratum sampling fraction to create population estimates. As discussed previously, approximately 1,250 students must be sampled in the higher-incidence disability categories, 1,012 students with autism, 559 students with traumatic brain injury, and 122 with deaf-blindness to make national estimates with reasonable precision about students in each category and students receiving special education overall.

Student Sample Selection Procedures

In Spring 2000, SRI will contact LEAs and obtain their agreement to participate in the study. In the fall of the 2000-2001 school year, we will request from participating LEAs rosters of students receiving special education between the ages of 13 and 16. Requests for rosters will specify that they contain identifiers for students receiving special education under the jurisdiction of the LEA, the disability category of each student, and the students' birthdates. As mentioned previously, some LEAs can be expected to provide only identification numbers for students, along with the corresponding birthdates and disability categories. When students are sampled in these LEAs, identification numbers of selected students are provided to the LEA, along with materials to mail to their parents/guardians (without revealing their identity to the study contractor).

After estimating the number of students receiving special education at the appropriate grade levels, we must determine the fraction of students in each category at each age that must be selected randomly from each district to yield a sample of 12,943 students. These sampling fractions will be calculated to maximize the effective sample efficiency while obtaining the required absolute sample sizes. Final sampling fractions cannot be calculated until the composition of the sample of participating LEAs is known; however, initial estimates are presented in Table 14.

Table 14				
ESTIMATED STUDENT SAMPLING FRACTIONS, BY LEA SIZE STRATUM				
(PERCENT): Age 13-15 /16				
	Very Large	Large	Medium	Small
Specific learning disability	0.37 / 0.56	0.49 / 0.74	1.54 / 2.31	4.61 / 9.92
Speech or language impairment	6.5 / 9.7	8.0 / 11.9	25.7 / 38.6	77.0 / 100
Mental retardation	1.8 / 2.7	2.2 / 3.3	7.4 / 11.1	22.3 / 33.4
Serious emotional disturbance	1.9 / 2.9	2.4 / 3.7	7.8 / 11.7	22.8 / 34.2
Multiple disabilities	15.6 / 23.5	20.1 / 30.2	64.0 / 96.0	100 / 100
Hearing impairments	18.6 / 27.8	23.9 / 35.8	77.5 / 100	100 / 100
Orthopedic impairments	26.6 / 39.9	35.0 / 52.5	100 / 100	100 / 100
Other health impairments	8.0 / 11.9	10.3 / 15.5	32.9 / 49.3	96.5 / 100
Visual impairments	50.0 / 75.0	59.0 / 100	100 / 100	100 / 100
Autism	100 / 100	100 / 100	100 / 100	100 / 100
Deaf-blindness	100 / 100	100 / 100	100 / 100	100 / 100
Traumatic brain injury	100 / 100	100 / 100	100 / 100	100 / 100

In addition, from the state-operated special schools that are recruited into the study, we will sample 100% of students who are deaf-blind, 100% of students with visual impairments, 20% of those age 13 to 15 with hearing impairments, and 30% of those age 16 with hearing impairments.

Student sampling weights are the product of the LEA sampling weights and the inverse of the student sampling fraction. The student sampling weight is the number of students in the universe represented by an individual student in the sample. Estimated sampling fractions and weights are included in Table 15. In addition, from the state-supported special schools, we expect sampling weights of 3.7 for the deaf-blind, 3.7 for students with visual impairments, 18.5 for students age 13 to 15 with hearing impairments, and 13.4 for students age 16 with hearing impairments.

Table 15				
EXPECTED STUDENT SAMPLING WEIGHTS. BY LEA SIZE STRATUM				
(Ages 13 to 15 /16)				
	Very Large	Large	Medium	Small
Specific learning disability	1,111 / 741	1,074 / 716	1,100 / 733	1,081 / 720
Speech or language impairment	64 / 43	66 / 44	66 / 44	65 / 50
Mental retardation	229 / 153	237 / 158	228 / 152	224 / 149
Serious emotional disturbance	216 / 144	217 / 144	217 / 145	219 / 146
Multiple disabilities	26 / 18	26 / 17	26 / 18	50 / 50
Hearing impairments	22 / 15	22 / 15	22 / 17	50 / 50
Orthopedic impairments	15.5 / 10.3	15.1 / 10.1	16.9 / 16.9	50 / 50
Other health impairments	52 / 35	51 / 34	51 / 34	52 / 50
Visual impairments	8.2 / 5.5	9.0 / 6.0	17 / 17	50 / 50
Autism	4.1 / 4.1	5.3 / 5.3	17 / 17	50 / 50
Deaf-blindness	4.1 / 4.1	5.3 / 5.3	17 / 17	50 / 50
Traumatic brain injury	4.1 / 4.1	5.3 / 5.3	17 / 17	50 / 50

Minimizing Sample Attrition

To minimize sample attrition over the years of data collection, the NLTS2 study contractor will need to use aggressive tracking mechanisms to maintain accurate and up-to-date contact information for sample members. To aid in this task, the parent questionnaire will include information that will facilitate tracking of parents/guardians, such as additional work and home telephone numbers for the respondents, location information for one or more friends or relatives who would know where the family had moved, and e-mail addresses.

REFERENCE

SRI International. (January 20, 2000). *The National Longitudinal Transition Study-2 (NLTS2) Conceptual Framework and Research Questions*. Menlo Park, CA: Author.