



The Study of State and Local Implementation and Impact of the  
**Individuals with Disabilities Education Act**



*Volume II:*

# **Technical Appendices**

## **Final Interim Report**

### **(1999-2000 School Year)**

March 2003

*Prepared for*  
Kelly Henderson, Ph. D.  
**U.S. Department of Education**  
Office of Special Education Program  
U.S. Department of Education  
Mary Switzer Building  
Room 4626  
330 C Street, S.W.  
Washington, DC 20202-2731

*Prepared by*  
Ellen Schiller  
Cristofer Price  
Patrick Johnston  
Amanda Parsad  
Kristine Burnaska  
Fran O'Reilly

Abt Associates Inc.



Abt Associates, Inc.  
4800 Montgomery Ln.  
Suite 600  
Bethesda, MD 20814-5341

**Volume II: Technical  
Appendices to Draft  
Final Interim Report—  
1999–2000 School Year**

March 2003

*Prepared for*  
Dr. Kelly Henderson  
U.S. Department of Education  
Office of Special Education Programs  
Mary Switzer Building, Rm. 4626  
330 C Street, SW  
Washington, DC 20202-2731

*Prepared by*  
Ellen Schiller  
Cris Price  
Patrick Johnston  
Amanda Parsad  
Kristine Burnaska  
Fran O'Reilly

Abt Associates Inc.



# Contents

<b>Appendix A. Design Objectives.....</b>	<b>A-1</b>
Describing Changes Over Time.....	A-1
Generalizability .....	A-1
Precision of Statistical Estimates.....	A-1
Subgroup Estimates.....	A-2
Contextual Information .....	A-2
Nonresponse Bias .....	A-3
<b>Appendix B. Selecting the Sample of Districts.....</b>	<b>B-1</b>
<b>Overview.....</b>	<b>B-1</b>
<b>The Sampling Frame of Districts for the Supplemental Sample .....</b>	<b>B-2</b>
The OCR Data .....	B-2
Deletions from the Preliminary List to Create the Sampling Frame .....	B-2
Region and Urbanicity Classifications .....	B-4
<b>Sampling Method.....</b>	<b>B-4</b>
<b>District Sample.....</b>	<b>B-4</b>
<b>Appendix C. Initial Sampling Weights for Districts .....</b>	<b>C-1</b>
<b>The Size of the Target Population.....</b>	<b>C-1</b>
Initial Sampling Weights for S1-Districts.....	C-2
Initial Sampling Weights for S2-Districts.....	C-3
Sum of the Weights of S1- and S2-Districts .....	C-6
<b>Appendix D. Selecting the Sample of Schools.....</b>	<b>D-1</b>
<b>Overview.....</b>	<b>D-1</b>
<b>The School Sampling Frame .....</b>	<b>D-2</b>
<b>Sampling Method.....</b>	<b>D-5</b>
<b>Description of the School Sample .....</b>	<b>D-7</b>
<b>Appendix E. Initial Sampling Weights for Schools.....</b>	<b>E-1</b>
<b>Overview.....</b>	<b>E-1</b>
<b>Weights for S1-Schools.....</b>	<b>E-1</b>
<b>Weights for S2-Schools.....</b>	<b>E-6</b>
<b>Appendix F. The District Nonresponse Study .....</b>	<b>F-1</b>
<b>Response Rates and the Potential for Bias .....</b>	<b>F-1</b>
<b>Districts That Responded to the Nonparticipation Survey.....</b>	<b>F-3</b>
<b>Estimation of the Bias in Survey Responses .....</b>	<b>F-4</b>
<b>Results: Estimated Bias for Each of the Nonresponse Survey Items .....</b>	<b>F-8</b>

<b>Appendix G. Recalibrating Weights to Reduce Nonresponse Bias .....</b>	<b>G-1</b>
<b>Starting Out: The First Recalibration.....</b>	<b>G-2</b>
<b>Subsequent Recalibrations .....</b>	<b>G-3</b>
<b>Recalibration Based on Multicategory Response Variables.....</b>	<b>G-3</b>
<b>Results of the Recalibration Process.....</b>	<b>G-4</b>
<b>A Note on Item Nonresponse.....</b>	<b>G-15</b>
<b>Appendix H. Items on the Nonresponse Surveys .....</b>	<b>H-1</b>
<b>Items on the District Nonresponse Survey.....</b>	<b>H-1</b>
<b>Items on the School Nonresponse Survey.....</b>	<b>H-5</b>
<b>Appendix I. Statistical Methods .....</b>	<b>I-1</b>
<b>Overview.....</b>	<b>I-1</b>
<b>Data .....</b>	<b>I-1</b>
Groups of Variables .....	I-1
School, District and State Outcomes.....	I-2
Continuous, Ordinal and Categorical Variables .....	I-3
Imputed Demographic Data .....	I-3
Sampling Zeros .....	I-4
<b>Logistic Models .....</b>	<b>I-5</b>
Model Selection.....	I-5

# Appendix A. Design Objectives

In Appendix A, we describe each design objective for the Study of State and Local Implementation and Impact of the Individuals with Disabilities Act (SLIIDEA) under two phases: design and implementation. The design phase began in September of 1998, when the U.S. Department of Education (ED) contracted with the American Institutes for Research (AIR) to design SLIIDEA. The implementation phase began on April 7, 2000, when Abt Associates Inc. and its subcontractors, Westat and SRI, were charged with collecting, analyzing and reporting on SLIIDEA data as articulated under the design phase. The objectives under the design phase were to

- describe changes over time in the implementation of the 1997 Amendments to IDEA at the state, district and school levels;
- collect data that are nationally generalizable;
- produce statistical estimates with acceptable levels of precision;
- produce statistical estimates for subgroups of interest; and
- provide qualitative contextual information to enrich the interpretation of nationally representative findings.

Under the implementation phase, one more design objective was included: to minimize potential nonresponse bias in study estimates. Each of the objectives is described below.

## Describing Changes Over Time

Changes over time can be measured in two ways: one is to collect data on different samples in successive years (cross-sectional samples), and the other is to collect data on the same sample across years (longitudinal samples). For the Core Survey, the design is such that data will be collected from the same longitudinal sample over time, in order to measure change in the intermediate and long-term outcomes of interest. For the Focus Study component, we will use a combination of cross-sectional and longitudinal samples. The sample selection and instrument design for both the Core Survey and the Focus Study components need to reflect issues in cross-sectional and longitudinal data collection and analysis. These include ensuring large enough sample sizes to account for attrition over time, and designing instruments that allow for analysis of items over time.

## Generalizability

The study findings must be generalizable to all school districts in the country and to all public schools. To meet this objective, the district and school samples needed to be selected in such a way that every school district in the country (and every school in the country) was represented in the study sample. The district sample becomes *nationally representative* by applying a sampling weight to each district, based upon each district's probability of being selected into the sample. Similarly, the school sample becomes *nationally representative* by applying a sampling weight to each school in the sample based upon each school's probability of being selected into the sample.

## Precision of Statistical Estimates

Precision refers to the width of the confidence intervals around the study estimates. The smaller the confidence intervals, the more precise the study estimates are, and thus, the more confidence we have in the study findings. Other things being equal, the larger the sample size, the narrower the width of

the confidence interval (i.e., the more precise the study estimates). It should also be noted that, with a given sample size, the confidence intervals are widest (i.e., the estimates are least precise) for study estimates of 50 percent. The width of the confidence intervals decreases as the study estimates get closer to 0 percent and 100 percent.

There is no universally acceptable level of precision; acceptable precision is subjective and is usually based on the consequences of the decisions being made using the study estimates. However, at some point, the confidence intervals might be so wide that the estimates are essentially meaningless. For this study, the expected level of precision is plus or minus four percentage points for districts, and two and a half percentage points for schools.

### Subgroup Estimates

In addition to producing reasonably precise estimates for districts and public schools, the study must be capable of producing reasonably precise estimates for subgroups of interest. In SLIIDEA, the important subgroups are school level—including elementary, middle and high schools—and student disability classification. During the design phase of SLIIDEA, 13 federal disability categories were clustered into four major disability categories: cognitive, behavioral, physical and sensory disabilities. (See Exhibit A1 for the four clusters.)

---

#### Exhibit A1

#### Clusters of Major Disability Categories Used in SLIIDEA

Disability Category			
Cognitive	Behavioral	Physical	Sensory
<ul style="list-style-type: none"> <li>• Specific learning disabilities</li> <li>• Speech or language impairments</li> <li>• Mental retardation</li> <li>• Developmental delay</li> </ul>	<ul style="list-style-type: none"> <li>• Emotional disturbance</li> <li>• Autism</li> </ul>	<ul style="list-style-type: none"> <li>• Orthopedic impairments</li> <li>• Other health impairments</li> <li>• Traumatic brain injury</li> <li>• Multiple disabilities</li> </ul>	<ul style="list-style-type: none"> <li>• Hearing impairments</li> <li>• Visual impairments</li> <li>• Deaf-blindness</li> </ul>

Any reasonably sized nationally representative sample of districts will almost certainly contain a nationally representative sample of schools that serves each school level and each of the major disability categories. The design issue is not the representativeness of the subgroup samples, but rather the size of the subgroup samples and the corresponding precision of the subgroup estimates. It should be noted that the precision of the subgroup estimates will always be less than that of estimates for the entire sample. For SLIIDEA, the expected level of precision for subgroup estimates is plus or minus five percentage points.

### Contextual Information

SLIIDEA is charged with not only obtaining quantitative data on a nationally representative sample of districts and schools, but also with collecting qualitative data in order to provide the overall study with rich contextualized information on how the translation of policy into practice occurs at the federal, state and district levels. This data will be collected and analyzed at state, district and school building levels, and will involve gathering information from a multiple of perspectives.

## **Nonresponse Bias**

Some sampled districts either have not or will not agree to participate in the study. This is problematic only to the extent that (1) nonresponders are different from responders in meaningful ways (i.e., in the outcomes of interest) **and** (2) the nonresponse rate is relatively high. Under these conditions the study estimates would be biased (i.e., the estimates would not be truly nationally representative). For example, it may not matter if nonresponders were larger or smaller; or more urban or less urban; or were more often located in the West rather than in another region. It would matter if nonresponders were implementing the 1997 Amendments of IDEA differently than responders. In general, the effects of potential nonresponse bias cause little concern if the nonresponse rate is less than 20 percent.



# Appendix B. Selecting the Sample of Districts

## Overview

In Appendix B, we describe the procedures for selecting the sample of districts. We begin this discussion with the design parameters. The SLIIDEA design documents produced by the American Institutes for Research (AIR) required a sample of 1,000 school districts and 3,800 schools. AIR had anticipated an 80 percent response rate for districts and schools and had therefore planned on conducting analyses based on 800 school districts. At the completion of the design phase, AIR had recruited only 324 districts to participate. The Office of Special Education Programs (OSEP) invited Abt to prepare several options for increasing the sample size.<sup>1</sup> OSEP selected an option whereby we would select a supplemental sample of districts with the intention of having a total of 800 recruited districts. We used the following strategy for obtaining 800 recruited districts:

- Three hundred and twenty-four districts had already been recruited. Therefore,  $(800-324=)$  476 additional districts needed to be recruited.
- To recruit 476 districts, the plan was to select a supplemental sample of 635 districts and assume that we would get a 75 percent agree-to-participate rate  $(635*0.75=476)$ .

Later in this document we describe the combined sample of the 324 districts that had already been recruited and the 635 districts selected for the supplemental sample. Thus, we address the selection of 959 districts in this appendix.

Selection of new districts involved the following considerations:

- If possible, the sampling frame for the supplemental sample of districts was to consist of those districts that had not already been selected as part of AIR's original sample. Districts that had been selected by AIR had already demonstrated their unwillingness to participate. The thinking was that it would be more difficult and far more expensive to convert districts that had already refused to participate into participating districts than it would be to entice new districts to participate.
- Districts that were selected by the SEELS and SPeNSE<sup>2</sup> studies were to be excluded from the sampling frame. It was believed that the response burden would be too great if a district that was already in SEELS or SPeNSE were asked to participate in SLIIDEA.
- Using National Center for Education Statistics (NCES) classifications, Type 1, 2 and 3 districts were to be included in the sampling frame. Types 4, 5, 6 and 7 were to be excluded. Exhibit B1 shows NCES codes for type of agency.

---

<sup>1</sup> Options were presented to OSEP in a June 20, 2000, document titled *Study of State and Local Implementation and Impact of the Individuals with Disabilities Education Act: Task 6.1.2 Alternative Sampling Plans*

<sup>2</sup> The SEELS and SPeNSE studies are two other OSEP-commissioned studies that are in the field. SEELS is the Special Education Elementary Longitudinal Study. SPeNSE is the Study of Personnel Needs in Special Education.

## The Sampling Frame of Districts for the Supplemental Sample

The sampling frame of districts was created from several sources. The preliminary list consisted of all records in the 1997–98 Common Core of Data (CCD). For reasons described below, we omitted many of the records from the preliminary list from the final sampling frame. Decisions about which records would be omitted and which would be retained in the sampling frame were based on information contained in the variables in

- the 1997–98 CCD,
- the 1997 and 1998 Office of Civil Rights (OCR) data sets,
- the list of the 1,004 districts that AIR had originally selected, and
- lists of districts in the SEELS and SPeNSE studies.

### The OCR Data

The samples of districts in each of the 1997 and 1998 OCR databases were about a third of the total number of districts in the country in each year. A district could have been selected to be in the OCR sample in any of six ways:

- 1) District was selected as part of the probability sample.
- 2) District was selected with certainty because it had 25,000 or more students.
- 3) District was selected with certainty because it was in a state where the total number of districts in the state was 25 or fewer (DC, DE, HI, MD, NV).
- 4) District was selected with certainty because it was under “court order/voluntary settlement.”
- 5) District was selected with certainty because it was in a “RESA <13” state.
- 6) District was selected with certainty because it had been a nonrespondent in two previous selection cycles.

Before the databases were merged to create the sampling frame, the districts that were included in the OCR samples for reasons 4, 5 and 6 were removed. We will refer to districts that entered the OCR sample for reasons 4, 5 or 6 as “forced-in” districts. The resulting samples included only those districts that had been selected because they were in the probability sample, were large (25,000+ students) or were in a state with 25 or fewer districts.

A single data file was created from the two databases (1997 and 1998). If a district had a record in both the 1997 and 1998 OCR databases, then the information from the more recent 1998 database was used.

### Deletions from the Preliminary List to Create the Sampling Frame

Districts in the preliminary list were omitted from the sampling frame for the following reasons:

- The preliminary list included districts from places that were not part of the 50 U.S. states or the District of Columbia. For example, districts in Puerto Rico, Guam and the Virgin Islands were removed from the list.
- Education agencies that had closed were removed from the list.

- Districts that had a grade span listed in the CCD as 0000 were removed from the list. There were only two such agencies that would not have been removed for other reasons. It is believed that the children in these districts are counted in other districts that remained in the list.
- Agencies that were regional, state, federal or other (nonlocal) were removed from the list. This information was in the TYPE97 variable on the CCD data set (see Exhibit B1).
- Districts with fewer than 100 students were removed from the list. We removed these very small districts because we believed it would be a waste of resources to recruit and mail to districts with so few students. Plenty of small districts remain in the sampling frame; they just had to have 100 or more students to be included.
- Districts that were in SEELS or SPeNSE were removed from the list.
- Our intention was to remove all of the districts that were in AIR’s original sample of 1,004 districts from the sampling frame. However, AIR’s original sample included all districts with more than 40,500 students. Some of these large districts had already been recruited and were included in the list of the 324 that had agreed to participate. Among the 324, there were three region-by-urbanicity strata that had very few of these large districts (West urban, West suburban, and Northeast urban). We decided that we would take a sample of nine large districts from the list of districts that had already been contacted by AIR. We therefore included large West urban, West suburban and Northeast urban districts that had already been selected by AIR in the sampling frame. Large districts from these three region-by-urbanicity strata were treated as separate strata for sampling purposes.
- Finally, after all the deletions described above, we calculated the number of new districts we needed to select from each of 12 region-by-urbanicity strata. We determined that the OCR databases contained a sufficient number of districts to allow us to restrict the sampling frame to only those districts that were in those databases. The reason for this restriction was that the information in OCR would be useful for selecting a sample of schools from the sample of districts. Sampling schools from OCR districts allows us to oversample schools that serve children with low-incidence disabilities.

<b>Exhibit B1</b>	
<b>Types of Districts Included and Excluded from Sampling Frame</b>	
	<b>NCES code for type of agency</b>
Included	1 = Local school district that is not a component of a supervisory union.
Included	2 = Local school district component of a supervisory union sharing a superintendent and administrative services with other local school districts.
Included	3 = Supervisory union administrative center, or a county superintendent serving the same purpose.
Excluded	4 = Regional education services agency, or a county superintendent serving the same purpose.
Excluded	5 = State-operated institution charged, at least in part, with providing elementary and/or secondary instruction or services to a special need population.
Excluded	6 = Federally operated institution charged, at least in part, with providing elementary and/or secondary instruction or services to a special need population.
Excluded	7 = Other education agencies that do not fit into the first six categories.

## Region and Urbanicity Classifications

All of the districts needed to be assigned to region and urbanicity classifications. Urbanicity was classified according to the metro status code variable on the CCD data set (see Exhibit B2). Region classifications were assigned based on the state that the district was located in. Classifications of states into the four census regions were obtained from a Census Bureau Web site (see Exhibit B3).

<b>Exhibit B2</b>	
<b>Urbanicity Classifications</b>	
<b>Classification</b>	<b>NCES classification of the agency's service area relative to a Metropolitan Statistical Area (CCD variable name MSC97)</b>
Urban	1 = Primarily serves a central city of an MSA
Suburban	2 = Serves an MSA but not primarily its central city
Rural	3 = Does not serve an MSA

<b>Exhibit B3</b>	
<b>Region Classifications</b>	
<b>Classification</b>	<b>State Abbreviation</b>
West	HI, WA, OR, MT, ID, WY, CA, NV, UT, CO, AZ, NM, AK
Northeast	NY, VT, ME, NH, MA, CT, RI, NJ, PA
South	TX, OK, AR, LA, KY, TN, MS, AL, WV, VA, MD, DC, DE, NC, SC, GA, FL
Midwest	ND, SD, NE, KS, MN, IA, MO, WI, IL, MI, IN, OH

## Sampling Method

We took a stratified systematic sample of 635 districts. Districts were sampled from 15 strata. The first 12 strata represented the region-by-urbanicity classifications and corresponded to districts that were not selected in AIR's original sample. The number of districts selected from these 12 strata was 626. A sample of nine districts was selected from three additional strata that consisted of large districts that had already been selected in AIR's original sample but had not been recruited.

For the 626 districts selected from 12 region-by-urbanicity strata, the number of districts to be sampled from within each of these strata was determined by proportional allocation. The allocation was proportional to the number of students<sup>3</sup>. Prior to sampling, we sorted the districts by size within each stratum. Systematic sampling after sorting by size increases the likelihood of having a wide distribution of district sizes in the selected sample.

<sup>3</sup> A simple example of proportional allocation is as follows. Suppose you had two strata (A and B) and were going to take a sample of 10 districts. Suppose there were 300 students in stratum A and 700 students in stratum B. We would then sample three districts from A and seven districts from B.

Exhibit B4 shows the total number of districts in the target, frame, sample and responder sample. Districts in the responder sample were those answering the survey, so these were the districts we were able to analyze. Contributions to the total are given separately based on the 333 S1-districts in the AIR district sample, and the 626 S2-districts in the Abt Associates district sample.

<b>Exhibit B4</b>			
<b>Number of Districts in the Population, Frame and Sample</b>			
	<b>S1-Districts</b>	<b>S2-Districts</b>	<b>Total</b>
Target Population	1,004	12,314	13,318
Sample Frame	1,004	11,190	12,194
Selected Sample	333	626	959
Responder Sample			298
Response Rate			31.07%

## District Sample

Exhibits B5, B6 and B7 show the (marginal) distributions for district urbanicity, region and size, respectively, for both the target population and the responder sample. Had allocation been proportional to the number of districts in the strata rather than to the number of students in a district, we would have expected similar distributions for the sample and the population for urbanicity, region and size. In fact, allocation was proportional to the number of *students* in the strata, and this has the effect of undersampling strata with many small districts, such as rural districts (Exhibit B5) and districts in the Midwest (Exhibit B6). As a consequence, smaller districts were undersampled and larger districts were oversampled (Exhibit B7).

<b>Exhibit B5</b>		
<b>Target and Sample by Urbanicity</b>		
	<b>Target Population</b>	<b>Responder Sample</b>
	Percent	
Rural	50.39	30.54
Suburban	44.00	53.69
Urban	5.61	15.77

<b>Exhibit B6</b>		
<b>Target and Sample by Region</b>		
	<b>Target Population</b>	<b>Responder Sample</b>
	Percent	
West	11.11	15.77
Northeast	15.51	16.78
South	29.86	36.58
Midwest	43.52	30.87

<b>Exhibit B7</b>		
<b>Target and Sample by District Size</b>		
	<b>Target Population</b>	<b>Responder Sample</b>
	Percent	
Small	63.55	39.51
Medium	28.44	42.95
Large	8.01	21.14

The hypothetical example below shows how this undersampling occurs. Suppose there are only two strata, each with 6,000 students, such that Strata 1 has 12 districts in the frame and Strata 2 has 6 districts in the frame. Thus Strata 1 has many small districts (500 students per district), and Strata 2 has few large districts (1,000 students per district). Because the number of districts in the sample is proportional to the number of students in the frame, equal numbers of districts are sampled from each strata, say three districts.

<b>Strata</b>	<b>Frame (Students)</b>	<b>Frame (Districts)</b>	<b>Sample (Districts)</b>
1	6,000	12	3
2	6,000	6	3
	12,000	18	6

The resulting proportions of districts in the two strata for the frame and the sample are shown below. Strata 1 has two-thirds of the districts in the frame but only half the districts in the sample, so Strata 1 is undersampled and Strata 2 is oversampled.

<b>Strata</b>	<b>Frame (Districts)</b>	<b>Sample (Districts)</b>
1	?	1/2
2	?	1/2

Exhibit B8 shows the joint distribution of district region, urbanicity and size for both the target population and the responder sample. The point here is to show that a sample of districts exists in all but a handful of cells where the target population exists. Exhibit B9 depicts the number of districts by state, again showing that the sample had good geographical coverage.

<b>Exhibit B8</b>				
<b>District Target and District Sample by Region, Urbanicity and District Size</b>				
<b>Region</b>	<b>Urbanicity</b>	<b>Size</b>	<b>Target Population</b>	<b>Responder Sample</b>
			Percent	
Midwest	Rural	Small	5.99	5.70
Midwest	Rural	Medium	6.22	3.36
Midwest	Rural	Large	0.36	0.00
Midwest	Suburban	Small	0.26	4.70
Midwest	Suburban	Medium	5.89	9.06
Midwest	Suburban	Large	0.98	2.35
Midwest	Urban	Small	0.87	0.00
Midwest	Urban	Medium	5.53	3.02
Midwest	Urban	Large	1.38	2.68
Northeast	Rural	Small	1.34	1.68
Northeast	Rural	Medium	3.93	0.67
Northeast	Rural	Large	0.38	0.00
Northeast	Suburban	Small	0.56	6.38
Northeast	Suburban	Medium	4.03	5.37
Northeast	Suburban	Large	1.20	0.00
Northeast	Urban	Small	0.69	0.00
Northeast	Urban	Medium	3.86	1.34
Northeast	Urban	Large	1.88	1.34
South	Rural	Small	4.70	6.71
South	Rural	Medium	8.02	7.38
South	Rural	Large	1.82	1.34
South	Suburban	Small	0.83	4.36
South	Suburban	Medium	4.76	5.70
South	Suburban	Large	3.20	5.70
South	Urban	Small	1.44	0.67
South	Urban	Medium	5.34	1.34
South	Urban	Large	4.76	3.36
West	Rural	Small	2.27	1.34
West	Rural	Medium	5.46	2.01
West	Rural	Large	0.51	0.34
West	Suburban	Small	0.29	3.36
West	Suburban	Medium	3.40	3.69
West	Suburban	Large	0.54	3.02
West	Urban	Small	1.50	1.01
West	Urban	Medium	3.91	0.00
West	Urban	Large	1.89	1.01

---

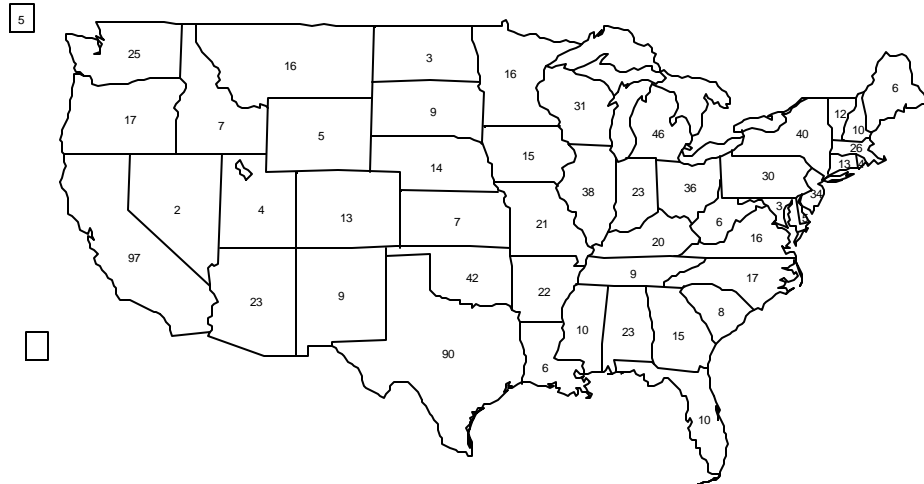
**Exhibit B9**

**Number of Districts Selected in Each State**

---

---

**Number of Districts Selected in Each State**



## Appendix C. Initial Sampling Weights for Districts

This appendix describes the calculation of initial sampling weights for districts. These weights would be used in the analyses only if there was no unit nonresponse (i.e., we obtained responses from all 959 districts in the sample). Subsequent sections of the appendices describe adjustments in the weights to deal with unit nonresponse.

The computation of the initial sampling weights is complex because of the way in which we arrived at our sample of 959 districts. In the design phase, the design contractor had selected 1,004 districts, of which they were able to recruit 324. We subsequently selected a supplemental sample of 635 districts, nine of which were intended to fill out some region-by-urbanicity cells that did not have enough large districts. As described earlier, the remaining 626 districts in the supplemental sample were selected from a sample of the districts in the combined 1997 and 1998 OCR database.

Definitions:

**S1-districts:** These are districts that are in the final sample and were also in AIR's original sample. There are 333 S1-districts. These include 324 districts that AIR recruited plus nine additional districts that we reselected, as described above, in order to fill out some region-by-urbanicity cells that did not have enough large districts.

**S2-districts:** These are districts that are in our supplemental sample of districts, but that were never part of the AIR sample. There are 626 S2-districts. These include the 635 districts selected in the supplemental sample, minus the nine large districts that were selected in the supplemental sample but had originally been part of AIR's sample.

### The Size of the Target Population

The size of the target population is 13,318 districts. The target population is the population of districts that we want to make inferences about (or, in other words, the population to which we want our sample to generalize). We are defining the target population to be broader than the sampling frame. This is because, for example, even though we excluded districts that were included in SEELS and SPeNSE from the sampling frame for the supplemental sample, we would like our results to generalize to the whole population that includes those districts.

The target population consists of the following:

- Type 1, 2 and 3 districts (i.e., excludes regional, state, federal or other (nonlocal) districts)
- Districts in the 50 states and the District of Columbia
- Districts that have schools and students associated with them (i.e., the number of schools in the districts is at least one and the grade span of students served exists)
- Districts that were not closed
- Districts with 100 or more students

The calculation of sampling weights is such that the sum of all the weights will be equal to 13,318 (the number of districts in the target population). One issue that makes the calculation of sampling weights more complex is that there were districts in SEELS and SPeNSE that were excluded from the

S1 and S2 sampling frames.<sup>4</sup> The rationale for excluding them was that we did not want to overburden those districts by selecting them for multiple studies. We are counting them as belonging to the target population of 13,318 even though they could not be in the sample. We are treating them as if they were missing completely at random.

Exhibit C1 shows some important numbers that will be referenced throughout this discussion. The first row shows that 11,190 districts comprised the sampling frame for the S2-districts. There were 1,124 districts that were excluded from S2 sampling frame because they were in SEELS or SPeNSE. There were 1,004 districts in the AIR sample. The cumulative sum of those three classifications of districts is 13,318. Those are the districts we are treating as our target population. There were 1,245 districts with fewer than 100 students that were excluded from the S2 sampling frame and are *not* considered to be part of the target population.

<b>Exhibit C1</b>				
<b>Size of Sampling Frame and Target Population</b>				
<b>Exclude</b>	<b>Frequency</b>	<b>Percent</b>	<b>Cumulative Frequency</b>	<b>Cumulative Percent</b>
Sample Frame for S2	11,190	76.84	11,190	76.84
SEELS/SPeNSE	1,124	7.72	12,314	84.56
In AIR sample	1,004	6.89	13,318	91.45
< 100 students <sup>a</sup>	1,245 <sup>a</sup>	8.55 <sup>a</sup>	14,563	100.00
<sup>a</sup> = Not in target population				

### Initial Sampling Weights for S1-Districts

Our approach is to treat AIR’s original sample of 1,004 districts as certainty units. That is, each district had a probability of selection equal to one. We then treat the 333 S1-districts as a stratified simple random sample of the 1,004 districts. The strata are the 12 region-by-urbanicity categories.

Sampling weights for S1-districts are calculated as follows. Let  $n_h$  be the population number of school districts (that were in AIR’s original sample) in the  $h^{th}$  stratum. Let  $m_h$  be the number of school districts in the  $h^{th}$  stratum that are in the S1 sample. Then the weight for the  $j^{th}$  district in the  $h^{th}$  stratum is calculated as:

$$w_{hj} = 1/\Pr(d_{hj} \in S1) = 1 / \frac{m_h}{n_h} = \frac{n_h}{m_h}. \quad \text{[Eqn 1.1]}$$

<sup>4</sup> Only those districts in SEELS and SPeNSE that were smaller than 40,500 were excluded. The SEELS and SPeNSE studies are other OSEP-commissioned studies that were being conducted at the same time as the SLIIDEA study.

Where  $\Pr(d_{hj} \in S1)$  is the probability that the  $j^{th}$  district in the  $h^{th}$  stratum is in sample S1. Note that

$$\sum_{h=1}^{12} n_h = 1,004, \sum_{h=1}^{12} m_h = 333, \text{ and } \sum_{h=1}^{12} \sum_{j \in S1} w_{hj} = 1,004.$$

The sum of the weights over all the districts in sample S1 is 1,004. In other words, these 333 districts represent 1,004 districts. Over all  $i$  districts in S1, we have:

$$\sum_{i \in S1} w_i = 1,004.$$

### Initial Sampling Weights for S2-Districts

As previously described, the S2-districts were selected from a sampling frame constructed from the 1997 and 1998 OCR databases. The combined OCR database was such that if a district was in either of the 1997 or 1998 samples, then it was included in the combined OCR sample (if a district was in both samples, then the record from 1998 was used).

The probability of selection of an S2-district is equal to the probability that the district was selected conditional on the district being in the combined OCR sample, multiplied by the probability that the district was in the combined OCR sample. We will use this notation:

$$\Pr(d_i \in S2) = \Pr(d_i \in S2 | d_i \in OCR) \times \Pr(d_i \in OCR),$$

[Eqn 1.2]

where  $d_i$  is district  $i$ .

Exhibit C2 shows the size of the sampling frame for S2-districts. The exhibit shows that in the first stage of sampling, 5,145 districts made it into the OCR sample. Note that in the second stage, a sample of districts was selected from those 5,145.

<b>Exhibit C2</b>					
<b>Size of Sampling Frame of S2-Districts</b>					
<b>Exclude</b>	<b>IN OCR</b>	<b>Frequency</b>	<b>Percent</b>	<b>Cumulative Frequency</b>	<b>Cumulative Percent</b>
Sample Frame for S2	.	6,045	54.02	6,045	54.02
Sample Frame for S2	1	5,145	45.98	11,190	100.00

### *The Conditional Probability of Selection*

The selection of the sample of S2-districts was described in previous sections of this appendix. We used SAS Proc Survey to select the sample of S2-districts from the sample of 5,145 OCR districts. The SAS procedure calculated probabilities of selection for each selected district. Thus, from the SAS output data set we have, for each S2-district, the probability of selection conditional on the district having been in OCR. That is, we know this part of the equation:

$$\Pr(d_i \in S2 | d_i \in OCR)$$

The result of summing the inverse of the conditional probability of selection over all of the districts in S2 is 5,145, as it should be. Now we turn to the problem of calculating the probability that a district was in OCR.

### ***The Probability of Selection into OCR***

For each district in the 1997 OCR database, we know its probability of selection. For each district in the 1998 OCR database, we know its probability of selection. We obtained these selection probabilities from the OCR contractor. In this section, we calculate the probability that a district was in the combined OCR sample. That is, we need to calculate this part of the equation:

$$\Pr(d_i \in OCR)$$

The logic here for getting the probability that a district was selected to be in either the 1997 or 1998 (or both) OCR databases is as follows:

Let  $\Pr(A)$  be the probability that a district was selected to be in the 1997 OCR database.  
Let  $\Pr(B)$  be the probability that a district was selected to be in the 1998 OCR database.  
Let  $\Pr(A \text{ or } B)$  be the probability that a district was in the combined OCR database.

Then,

$$\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B)$$

Selection in 1997 was independent of selection in 1998. Therefore,  
 $\Pr(A \text{ and } B) = \Pr(A) * \Pr(B)$ .

Now, consider a district that is in the combined OCR database. That is, the district is in A or B. Suppose that the district was in A, but not B. Then we know  $\Pr(A)$  because it was attached to the file we obtained from the OCR contractor, but we do not know  $\Pr(B)$ . Therefore, we can't calculate  $\Pr(A \text{ or } B)$  as shown above. We do know that the sampling strategies were very similar for the 1997 and 1998 samples. Therefore, we used the following method to approximate the unknown  $\Pr(B)$ : If we don't know  $\Pr(B)$  for a district (its probability of selection in 1998), we use  $\Pr(A)$  in its place (its probability of selection in 1997). Our assumption is that a district's probability of selection in 1998 will be very similar to its probability of selection in 1997. Likewise, if we know  $\Pr(B)$  but don't know A, then we approximate  $\Pr(A)$  by  $\Pr(B)$ .

How good are our estimated probabilities of selection? The sum of the inverse of the probabilities of selection into OCR over the 5,145 districts that were in OCR and in the S2 sampling frame is 10,024. We would expect the sum of these weights to be close to the number shown in the first row of Exhibit B9. That number is 11,190. These two numbers are not closer for two reasons. The first is that our approximation method might be off somewhat. The other is that there was some nonresponse in the OCR sample. Not every district that was selected to be in the 1997 OCR sample responded, nor did every district in the 1998 sample respond.

Our approach was to adjust the probabilities of selection into OCR so that the sum of their inverses across the 5,145 districts would be 11,190. We calculated those adjustments within the 12 region-by-urbanicity categories.

Let  $w_{hj}^{adj}$  be the new weight for the  $j^{th}$  district in the  $h^{th}$  stratum after the adjustment. And let  $w_{hj}$  be the weight before adjustment. These weights are the inverse of the approximated probability of selection into OCR that has been calculated up to this point (i.e., the weights that sum to 10,024). Let  $n_h$  be the population number of school districts in the  $h^{th}$  stratum. Let  $m_h$  be the number of school districts in the sample. Then the new adjusted weight for the  $j^{th}$  district in the  $h^{th}$  stratum is calculated as:

$$w_{hj}^{adj} = w_{hj} \times \frac{n_h}{\sum_{j=1}^{m_h} w_{hj}} . \quad \text{[Eqn 1.3]}$$

Note that  $\sum_{h=1}^{12} n_h = 11,190$  ,  $\sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj} = 10,024$  , and  $\sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj}^{adj} = 11,190$ .

Let new, adjusted probabilities be denoted:

$$\Pr(d_i \in OCR) = 1/w_i^{adj} ,$$

where the  $i$  subscript corresponds to all the districts in all strata.

### ***The Unconditional Probability of Selection***

We now have both parts of the right-hand side of Eqn 1.2, and can therefore calculate, for each district in the S2 sample, the left-hand side:

$$\Pr(d_i \in S2) .$$

As usual, the weights will be equal to the inverse of the probability of selection,

$$w_i^* = 1 / \Pr(d_i \in S2) . \quad \text{[Eqn 1.4]}$$

The weights in Equation 1.4 are starred (“\*”) to differentiate them from the adjusted weights described in the next section. Summing the weights over all districts in S2 gives

$$\sum_{i \in S2} w_i^* = 11,163.73$$

This is pretty close to the population size of 11,190 (shown in the first row of Exhibit A9).

### ***Rewighting so That S2-Districts Will Represent the Population That Includes SEELS and SPeNSE Districts***

We now adjust the weights shown in Eqn. 1.4 so that they sum to 12,314. This is the number of districts in the S2 sampling frame plus the districts that were excluded because they were in SEELS or SPeNSE (see Exhibit B1). The rationale for this adjustment is that we believe that it is reasonable to treat the S2 sample of districts as representative of the population that includes the SEELS and SPeNSE districts. If we can treat the sample as if it represents the full population including the SEELS and SPeNSE districts, then we avoid repeated explanations in the results sections of our reports that the SEELS and SPeNSE districts were excluded from the sampling frame.

The weighting adjustment would use the same method as described in Eqn 1.3. After adjustment, summing the new weights over all districts in S2 gives

$$\sum_{i \in S2} w_i = 12,314.$$

### **Sum of the Weights of S1- and S2-Districts**

Summing over the weights of all 959 districts in the sample gives

$$\sum_{i \in S1 \cup S2} w_i = \sum_{i \in S1} w_i + \sum_{i \in S2} w_i = 1,004 + 12,314 = 13,318.$$

These weights represent the initial sampling weights. Subsequent adjustments are necessary to adjust for unit nonresponse. These adjustments are described in Appendix G.

# Appendix D. Selecting the Sample of Schools

## Overview

### *Sampled Schools Within Sampled Districts*

This appendix describes the procedures for selecting the sample of schools. Recall that the selection of districts was based on a systematic sample (based on district size) within a stratified sample (based on the 12 region-by-urbanicity strata). Both types of sampling had the effect of increasing the spread of districts across all combinations of region, urbanicity and size. This first stage of the sampling plan is described in detail in Appendix B.

The second stage of the sampling plan consisted of sampling schools from the districts selected at stage one. This was done separately for each school level (elementary, middle and high) to enable separate analyses for these subpopulations. Thus sampled schools were clustered (or nested) within sampled districts.

Various considerations affected the way in which schools were selected. First, we wanted the sample of schools to be representative of all public schools in the country. Second, we wanted to select a sample of elementary, middle and high schools from each of the selected districts. Third, we wanted to oversample schools serving large numbers of children with sensory and physical disabilities because of the relatively low prevalence of such schools. Finally, we wanted to avoid enormous variation in school selection probabilities because this could lead to very large variance estimates, and thus very wide confidence intervals.

We originally intended to sample two elementary schools, two middle schools and two high schools from each district, with an emphasis on sampling schools with large numbers of sensory and physically disabled students. When sampling two or more schools of the same type (e.g., two elementary schools with similar numbers of sensory and physically disabled students), we sampled systematically by school size. The two-stage cluster sampling plan for a given school level (e.g., elementary schools) is depicted below.

Stage	Sampling Unit	Stratification Factors	Systematic Factors	Oversampling Factors
1	district	region, urbanicity	district size	—
2	school		school size	sensory and physically disabled students

## The School Sampling Frame

### *Frame and Target*

Recall from Appendix B that a total of 959 districts were selected in the first stage of the sample. This was the sum of the 333 S1-districts selected from the AIR district sampling frame, and the 626 S2-districts selected from the Abt Associates district sampling frame. The school sampling frame comprised the 13,900 schools that were in the 959 selected districts and in the 1997–98 NCES Common Core of Data (CCD), minus 665 schools that were considered inappropriate for reasons discussed below. Thus the school sampling frame consisted of all 13,235 appropriate schools from the 959 sampled districts.

The school *sampling frame* is the list of schools from which the sample was actually drawn, and the school *target population* is the list of schools that the sample is supposed to represent. Because selected schools were nested in selected districts, the school target population is far larger than the school sampling frame. Ignoring minor exceptions, the school sampling frame consisted of all 13,900 CCD schools in the 959 sampled districts, while the school target population consisted of all 82,947 CCD schools in all 13,318 districts. In other words, the school target population consisted of all CCD schools in the country (defined as the 50 states and DC).

### *Exclusions from the Frame*

We now discuss the 665 schools that were excluded from the school sampling frame. The sampling frame originally included 13,900 schools. We omitted 129 schools from the sampling frame because they were closed. We omitted 150 because they served only kindergarten or prekindergarten children (they did not have any students in grade ranges of 1 to 12). We omitted 153 schools because they were very small (15 or fewer students). And 233 schools were omitted because they were listed in the CCD database as “grade span=0000.” For these schools, we had no information on the number of children that were served or the grade levels of children that were served in the school. The personnel providing data for the CCD were instructed not to count any child twice if they were simultaneously enrolled in two schools. Typically, students in “grade span=0000” schools are counted in their “home schools.” Thus children in these schools should be represented in the sampling frame because their home schools are represented in the sampling frame. We cannot be sure of this, however; we simply do not know very much about these 233 schools. After these omissions, the sampling frame consisted of 13,235 schools.

### *Target, Frame and Sample: Elementary, Middle and High Schools*

As mentioned previously, the sample was not actually drawn from the All Schools Combined frame (13,235 schools), but rather from three distinct frames: schools teaching elementary students (8,853 schools), schools teaching middle students (3,763 schools) and schools teaching high school students (2580 schools). Exhibits D1 through D3 show the numbers of schools in the target, frame and sample for schools teaching elementary, middle and high school students. They also show the number of schools that received surveys in the mail. The data collection protocol required active consent from each school’s district before a survey could be mailed to that school. When consent was not obtained, the survey was not mailed to the school. The tables also show the number of schools that returned completed questionnaires. These responder schools are the ones we were able to analyze. Two response rates are shown in each table. The first shows the responder schools as a percentage of the number of selected schools. The second shows the responder schools as a percentage of the number of schools that received a survey in the mail, and therefore had an opportunity to respond.

<b>Exhibit D1</b>			
<b>Elementary Schools</b>			
	<b>S1-Schools</b>	<b>S2-Schools</b>	<b>Total</b>
A) Target Population	18,320	34,360	52,680
B) Sample Frame	6,306	2,547	8,853
C) Selected Sample	905	1,272	2,177
D) Consent Obtained and Mailed to			904
E) Responder Sample			377
F) Response Rate 1 (100*E/C)			17%
G) Response Rate 2 (100*E/D)			42%

<b>Exhibit D2</b>			
<b>Middle Schools</b>			
	<b>S1-Schools</b>	<b>S2-Schools</b>	<b>Total</b>
A) Target Population	7,221	17,450	24,671
B) Sample Frame	2,692	1,071	3,763
C) Selected Sample	674	884	1,558
D) Consent Obtained and Mailed to			724
E) Responder Sample			281
F) Response Rate 1 (100*E/C)			18%
G) Response Rate 2 (100*E/D)			39%

<b>Exhibit D3</b>			
<b>High Schools</b>			
	<b>S1-Schools</b>	<b>S2-Schools</b>	<b>Total</b>
A) Target Population	5,263	13,821	19,084
B) Sample Frame	1,787	793	2,580
C) Selected Sample	589	678	1,267
D) Consent Obtained and Mailed to			619
E) Responder Sample			206
F) Response Rate 1 (100*E/C)			16%
G) Response Rate 2 (100*E/D)			33%

**Target, Frame and Sample: All Schools Combined**

Exhibit D4 shows the number of schools in the frame and sample for All Schools Combined. The sampling frame for All Schools Combined (13,235 schools) was less than the sum of the three separate elementary, middle and high school frames (8,853 + 3,763 + 2,580 = 15,196 schools) because a school teaching both elementary and middle students, for example, is in both the elementary and middle frames, but appears only once in the All Schools Combined frame. Similarly, the sample for All Schools Combined (4,534 schools) was less than the sum of the three elementary, middle and high schools samples (2,177 + 1,558 + 1,267 = 5,002 schools). This is because some schools teaching multiple levels were also sampled for multiple levels.

<b>Exhibit D4</b>			
<b>Multiple School Levels</b>			
<b>Type of School</b>	<b>Sampling Frame (1)</b>	<b>Sample – Levels Taught (2)</b>	<b>Sample – Levels Selected (3)</b>
Elementary School only	7,617	1,901	1,954
Middle School only	2,096	1,026	1,157
High School only	1,855	946	1,016
Elementary and Middle School	942	257	156
Elementary and High School	0	0	6
Middle and High School	431	236	184
Elementary, Middle and High School	294	168	61
Total	13,235	4,534	4,534

However, not all schools teaching multiple levels were selected for all of their levels. This distinction appears in Exhibit D4 as sampled schools *in terms of the levels taught* (column two) and sampled schools *in terms of the levels sampled* (column three). For example, if a school teaching both elementary and middle students is sampled, it need not be in both the elementary and middle sample. Rather it will be in either the elementary-only sample, the middle-only sample or the elementary-and-middle sample. Suppose the school was selected for the elementary-only sample. Then in terms of Exhibit D4, the school would appear on the “Elementary and Middle School” row for the sampling frame (column one), and for the sample in terms of levels taught (column two), but would appear on the “Elementary School Only” row for the sample in terms of levels sampled (column three).

In general then, the fact that some schools teaching multiple levels were not selected for all of their levels makes some of the schools in the lower rows of column three migrate upward to form column four. This explains the apparent incongruity whereby six schools were in the elementary-and-high-school sample, even though no such schools existed in the sampling frame. These schools actually taught all three levels, but were selected in the elementary and high school samples and were not selected for the middle school sample. Hence, these schools appear in the “Elementary, Middle and High School” row of column two, and the “Elementary and High School” row of column three.

Exhibit D5 shows, for All Schools Combined, the total number of schools in the target, frame, sample and responder sample. Contributions to the total are given separately based on the 333 S1-districts in the AIR district sample, and the 626 S2-districts in the Abt Associates district sample.

<b>Exhibit D5</b>			
<b>All Schools</b>			
	<b>S1-Schools</b>	<b>S2-Schools</b>	<b>Total</b>
A) Target Population	27,141	55,806	82,947
B) Sample Frame	9,295	3,940	13,235
C) Selected Sample	1,978	2,556	4,534
D) Consent Obtained and Mailed to			1,897
E) Responder Sample			744
F) Response Rate 1 (100*E/C)			16%
G) Response Rate 2 (100*E/D)			39%

## Sampling Method

The sample for All Schools Combined can be broadly construed as a two-stage cluster design, with stratification, systematic sampling and oversampling as depicted below.

<b>Stage</b>	<b>Sampling Unit</b>	<b>Stratification Factors</b>	<b>Systematic Factors</b>	<b>Oversampling Factors</b>
1	district	region, urbanicity	district size	—
2	school	school level	school size	sensory and physically disabled students

The stratification by school level should not be taken literally, however, because schools teaching multiple levels (13 percent of all schools) appeared in multiple sampling frames. In fact, the second stage of the sampling plan was executed separately for each school level, based on the following two-stage cluster design.

<b>Stage</b>	<b>Sampling Unit</b>	<b>Stratification Factors</b>	<b>Systematic Factors</b>	<b>Oversampling Factors</b>
1	district	region, urbanicity	district size	—
2	school		school size	sensory and physically disabled students

### ***Sampling School Levels***

Our original intention was to select two schools from each level (two elementary, two middle, two high) from each district selected in the first stage of the sample. In fact, many districts had too few schools to make this possible, and the sampling plan evolved accordingly. The typical situation was one in which many districts had enough elementary schools but too few middle and high schools. We sampled what was possible, but this left an overall shortfall of schools in the sample. For example, if a district had four elementary, one middle and one high school, we sampled two elementary, one middle and one high school, leaving a deficit of two schools. As the sample proceeded in this fashion, we needed to increase the desired sample per district to, say, five schools from each level. The final result was that we obtained more elementary schools than middle and high schools: of the 4,534 selected schools, 2,177 were elementary schools, 1,558 were middle schools and 1,267 were high schools.

This imbalance occurred, of course, because elementary schools are far more numerous (in the frame and the population) than middle and high schools (see Exhibits D1 through D3). The effect was to redistribute the sample of 5,002 schools (2,177 elementary, 1,558 middle, 1,267 high) from an intended design of equal allocation (1,668 elementary, 1,667 middle, 1,667 high, say) toward a design of proportional allocation (2,914 elementary, 1,239 middle, 849 high). The common proportion here is 0.33, being the sum of the samples (5,002) divided by the sum of the frames (15,196).

### ***Oversampling Sensory and Physical Disabilities***

Thirty-nine percent of schools in the sampling frame served children with sensory disabilities. The percentages for physical, behavioral and cognitive disabilities were 62 percent, 70 percent and 92 percent, respectively. It seemed sensible therefore, to oversample schools serving children with sensory and physical disabilities. Schools serving children with cognitive and behavioral disabilities were not oversampled because these types of schools were more prevalent.

The details of the oversampling algorithm are complicated, but the main point is that whenever more than two schools of a given level (elementary, say) existed in a district, schools with high numbers (10 or more students) of sensory disabled students and schools with high numbers of physically disabled students were given preference in the selection process. For example, if a district had 20 elementary schools—one with many sensory cases, one with many physical cases and 18 “other” schools—and we wanted to draw three of these schools, then the algorithm would have given us one of each school type (one with many sensory cases, one with many physical cases and one “other”).

### ***Sampling School Size***

In situations where one or more schools of a given level were to be selected from a larger group of a given type, schools were selected systematically by school size using Lahari’s circular method. This has the effect of ensuring a reasonable spread of school sizes in such cases. For example, suppose we wanted to draw a sample of three from seven similar elementary schools (e.g., all with less than 10 sensory and physical cases). We would rank the schools in order of size (1, 2, 3, 4, 5, 6, 7) pick a random start (2, say), and then choose every other school (4 and 6).

## Description of the School Sample

### *School Levels*

As previously noted, we intended to select two elementary, two middle and two high schools from each selected district, but many districts had too few schools to make this possible, and so we selected many more elementary schools than middle and high schools. We showed that the effect was to redistribute the sample from a design of equal allocation toward a design of proportional allocation. As Exhibit D7 shows, this effect carried over to the target population and the responder sample. The percent of elementary, middle and high schools in the responder sample is in between proportion allocation (the percentages in the target population) and equality (33 percent in each level). We remark here that each type of design—proportion allocation and equal allocation—has its advantages. Proportional allocation is more representative, while equal allocation is more efficient in certain situations (for example, it gives more precise estimates of the difference between two school-level means).

<b>Exhibit D7</b>		
<b>Target and Sample by School Level</b>		
	<b>Target Population</b>	<b>Responder Sample</b>
	Percent	
Elementary	63.95	50.67
Middle	31.01	37.77
High	19.56	27.69

### *Oversampling Sensory and Physical Disabilities*

We define the *service rate* of a particular disability as the percent of schools serving at least one student with that disability. We have noted previously that the service rates for sensory disabilities (39 percent) and physical disabilities (62 percent) were lower than those for behavioral disabilities (70 percent) and cognitive disabilities (92 percent). We oversampled schools serving children with sensory and physical disabilities to overcome this imbalance. The effect of such oversampling is unlikely to be dramatic because it could effectively operate only in large districts, and many districts in our sample were small. Nevertheless, as Exhibit D8 shows, the overall effect was successful because the percent of schools serving sensory and physical disabilities was higher in the sample than in the frame. This result applied to the three separate samples of elementary, middle and high schools (not shown), and Exhibit D8 was constructed from a weighted average of these tables.

<b>Exhibit D8</b>		
<b>Percent of Schools Serving Various Disability Categories: Sample Frame and Selected Sample—All Schools Combined</b>		
<b>Disability Served</b>	<b>Sample Frame</b>	<b>Selected Sample</b>
	Percent	
Sensory Impaired	39	45
Physical Disabilities	62	69
Cognitive Disabilities	92	92
Behavior Disabilities	70	70

**School Size, Region and Urbanicity**

Exhibits D9, D10 and D11 show the (marginal) distributions of school size, region and urbanicity for both the target population and the responder sample. We do not expect any particular relationship between the target population and the responder sample. In particular, we do not expect proportionality, because the original design would have made the proportion of selected schools in a region equal to the proportion of *districts* in the region, rather than equal to the proportion of the school *population* in the region. And indeed, this is exactly what we find, albeit in an attenuated form because the original equal allocation design moved toward a proportion allocation design as discussed previously. In the case of school size, the attenuation is yet stronger because of the systematic sampling by school size.

<b>Exhibit D9</b>		
<b>School Target and School Sample by School Size</b>		
	<b>Target Population</b>	<b>Responder Sample</b>
	Percent	
Small	20.75	16.13
Medium	60.35	54.70
Large	18.90	29.17

<b>Exhibit D10</b>		
<b>School Target and School Sample by Region</b>		
	<b>Target Population</b>	<b>Responder Sample</b>
	Percent	
Midwest	27.47	28.23
Northeast	17.87	15.19
South	34.87	40.59
West	19.78	15.99

<b>Exhibit D11</b>		
<b>School Target and School Sample by Urbanicity</b>		
	<b>Target Population</b>	<b>Responder Sample</b>
	Percent	
Rural	41.01	33.60
Suburban	25.93	31.72
Urban	33.06	34.68

Exhibit D12 shows the joint distribution of region, urbanicity and school size for both the target population and the responder sample. The point here is not to demonstrate proportionality (the departure from which is dealt with by the sampling weights in any case), but rather to show that a sample of schools exists in all cells where the target population exists—that is, in all cells.

**Exhibit D12****School Target and School Sample by Region, Urbanicity and School Size**

Region	Urbanicity	Size	Target Population	Responder Sample
			Percent	
Midwest	Rural	Small	5.99	3.23
Midwest	Rural	Medium	6.22	4.57
Midwest	Rural	Large	0.36	0.94
Midwest	Suburban	Small	0.26	0.27
Midwest	Suburban	Medium	5.89	6.32
Midwest	Suburban	Large	0.98	3.09
Midwest	Urban	Small	0.87	1.61
Midwest	Urban	Medium	5.53	4.97
Midwest	Urban	Large	1.38	3.23
Northeast	Rural	Small	1.34	0.67
Northeast	Rural	Medium	3.93	3.36
Northeast	Rural	Large	0.38	0.54
Northeast	Suburban	Small	0.56	0.54
Northeast	Suburban	Medium	4.03	4.57
Northeast	Suburban	Large	1.20	1.21
Northeast	Urban	Small	0.69	0.81
Northeast	Urban	Medium	3.86	2.15
Northeast	Urban	Large	1.88	1.34
South	Rural	Small	4.70	2.42
South	Rural	Medium	8.02	10.22
South	Rural	Large	1.82	2.69
South	Suburban	Small	0.83	0.81
South	Suburban	Medium	4.76	4.97
South	Suburban	Large	3.20	5.24
South	Urban	Small	1.44	2.42
South	Urban	Medium	5.34	6.05
South	Urban	Large	4.76	5.78
West	Rural	Small	2.27	1.21
West	Rural	Medium	5.46	2.96
West	Rural	Large	0.51	0.81
West	Suburban	Small	0.29	0.67
West	Suburban	Medium	3.40	2.42
West	Suburban	Large	0.54	1.61
West	Urban	Small	1.50	1.48
West	Urban	Medium	3.91	2.15
West	Urban	Large	1.89	2.69



# Appendix E. Initial Sampling Weights for Schools

## Overview

This appendix describes the calculation of the initial set of weights for schools. These weights are appropriate assuming all schools in the sample respond to the survey. In cases where unit nonresponse occurs, weights were inflated accordingly. For example, if half the elementary schools in the urban West district did not respond, then the weights of the responding elementary schools in the urban West would double.

### *S1-Districts and Schools, and S2-Districts and Schools*

Recall from Appendix B that a total of 959 districts were selected in the first stage of the sample. This was the sum of the 333 S1-districts selected from the AIR district sampling frame of 1,004 districts, and the 626 S2-districts selected from the Abt Associates district sampling frame of 11,190 districts. S1-schools correspond to the sample of S1-districts, and S2-schools correspond to the sample of S2-districts.

The 333 S1-districts included the 324 districts selected by AIR plus nine additional districts selected by Abt Associates from AIR's original sampling frame of 1,004 districts. The nine additional districts were selected in order to fill out region-by-urbanicity cells that did not have enough large districts. The 626 S2-districts included the 635 districts selected in the supplemental sample, minus the nine large districts that were selected in the supplemental sample that were originally part of AIR's sampling frame.

S1-districts were weighted to represent the S1-target population of 1,004 districts, and S2-districts were weighted to represent the S2-target population of 12,314 districts. The sum of these weights is 13,318, the size of the total district target population.

## Weights for S1-Schools

### *Conditional Weights for S1-Schools*

A conditional weight is the inverse of the probability of selection of a school *conditional on* the school's district having been selected to be in the sample. The conditional weights were calculated in two separate SAS programs. The first program calculated weights for the schools selected from the 324 districts recruited by AIR. The second program calculated the weights for schools selected from the nine additional districts that were reselected in order to fill out region-by-urbanicity cells that did not have enough large districts.

Of the 324 districts recruited by AIR, recall that three independent samples of schools were taken: a sample of elementary schools, a sample of middle schools and a sample of high schools. A sample of 662 elementary schools was selected from a sampling frame of 4,562 elementary schools. The conditional weights for those 662 sampled elementary schools were readily obtained from the output from the SAS program used to select the sample. The sum of the sampling weights for those 662 schools is 4,562, as it should be. The conditional weights for middle and high schools were obtained the same way. The numbers of schools in the sampling frames, the numbers of schools selected and

the sums of their conditional weights for elementary, middle and high schools are summarized in Exhibit E1.

<b>Exhibit E1</b>		
<b>Schools Selected from 324 Districts Recruited by AIR</b>		
<b>School Level</b>	<b>Sums of Conditional Sampling Weights: Schools in the Sampling Frame</b>	<b>Schools Selected</b>
Elementary	4,562	662
Middle	2,001	532
High	1,261	498

In addition to conducting analyses at the school level, we also wished to run analyses for all schools combined. This requires the calculation of the probability that a school was selected for *any* of the three samples, namely

$$\begin{aligned} \Pr(E \text{ or } M \text{ or } H) = & \Pr(E) + \Pr(M) + \Pr(H) \\ & - \Pr(E \text{ and } M) - \Pr(E \text{ and } H) - \Pr(M \text{ and } H) \\ & + \Pr(E \text{ and } M \text{ and } H). \end{aligned}$$

Because the three samples were independent, all joint probabilities were simply the products of the marginal probabilities. Thus  $\Pr(E \text{ and } M) = \Pr(E) \cdot \Pr(M)$ ,  $\Pr(E \text{ and } H) = \Pr(E) \cdot \Pr(H)$ ,  $\Pr(M \text{ and } H) = \Pr(M) \cdot \Pr(H)$ , and  $\Pr(E \text{ and } M \text{ and } H) = \Pr(E) \cdot \Pr(M) \cdot \Pr(H)$ .

Combining the sampling frames for elementary, middle and high schools in the 324 districts recruited by AIR gave 6,729 unique schools. A *unique school* is a school selected into at least one of the elementary, middle or high school samples. In this context, a school is counted only once even if it was selected for more than one sample; for example, a school teaching elementary and middle students could be selected for both elementary and middle school samples. There were a total of 1,579 unique schools selected in the elementary, middle or high school samples. The program calculated the conditional probabilities of selection,  $\Pr(E \text{ or } M \text{ or } H)$ , and the initial weights (the inverse of these selection probabilities). The calculations were then summed over the 1,579 unique schools. The sum of the weights was 6,815.25, which was too high by 86.25, since the 1,579 schools in the sample are supposed to represent the 6,729 schools in the combined sampling frame. The weights were correspondingly adjusted to equal 6,729.

After adjusting the weights, we ensured that no school had a weight of less than one. Out of the 1,579, there were 666 representing themselves only (had sampling weights of one) and 913 purporting to represent the remaining 6,063 schools ( $6,729 - 666 = 6,063$ ). However, the sum of the weights for the 913 schools was 6,149.25 (86.25 too many), so the weights were reduced by a factor of  $6,063/6,149.25$  for these 913 schools. The new adjusted weights for the 1,579 schools summed to 6,729, as they should.

---

**Exhibit E2****Schools Selected from 324 Districts Recruited by AIR**

School Level	Sums of Conditional Sampling Weights: Schools in the Sampling Frame	Schools Selected
All Schools Combined	6,729	1,579

The logic applied to the calculation of weights for schools selected from the 324 districts was applied to the calculation of weights for schools selected from the nine supplemental districts.

---

**Exhibit E3****Schools Selected from Nine Supplemental Districts**

School Level	Sums of Conditional Sampling Weights: Schools in the Sampling Frame	Schools Selected
Elementary	1,744	243
Middle	691	142
High	526	91
All Schools Combined	2,566	399

Combining results from both programs, we obtain weights for all of the schools selected from S1-districts, as shown in Exhibit E4.

---

**Exhibit E4****Schools Selected from All 333 S1-Districts (the 324 and the 9)**

School Level	Sums of Conditional Sampling Weights: Schools in the Sampling Frame	Schools Selected
Elementary	6,306	905
Middle	2,692	674
High	1,787	589
All Schools Combined	9,295	1,978

***Unconditional Weights for S1-Schools***

An unconditional weight is the inverse of the unconditional probability of selection, which is the school's conditional probability of selection multiplied by its district's probability of selection:

$$\Pr(s_{ij} \in S1) = \Pr(s_{ij} \in S1 | d_i \in S1) \times \Pr(d_i \in S1)$$

where  $d_i$  is district  $i$ , and  $s_{ij}$  is school  $j$  in district  $i$ .

The weight variable associated with the probability that a district was selected is called “wgt5”. Note that the sum of  $wgt5$  for S1-districts is 1004 because the 333 S1-districts represent the 1,004 districts in the AIR sampling frame.

The conditional probability of selection, specified as:

$$\Pr(s_{ij} \in S1 | d_i \in S1)$$

is the inverse of the adjusted weight for the All Schools Combined weights ( $allwgt2$ ). These are simply the selection probabilities output by SAS in the case of elementary, middle and high school samples.

The program calculates the unconditional probability of selection,

$$\Pr(s_{ij} \in S1)$$

via the statements:  $scprb1=(1/wgt5)*(1/allwgt2)$ ,  $Elemprb1=(1/wgt5)*elemSelectionProb$ ,  $Midlprb1=(1/wgt5)*midlSelectionProb$ , and  $Highprb1=(1/wgt5)*highSelectionProb$ . As usual, the weights are the inverses of the corresponding probabilities. Exhibit 5 shows the target population of schools in the 1,004 districts; the sample of S1-schools is intended to represent this population.

<b>Exhibit E5</b>	
<b>Numbers of Schools in the 1,004 Districts</b>	
<b>School Level</b>	<b>Number of Schools</b>
Unique Schools	27,141
Elementary Schools	18,320
Middle Schools	7,221
High Schools	5,263

However, Exhibit E6 shows that the sum of the weights from the sampled schools was slightly higher than the number of schools in the 1,004 districts. School weights were adjusted in order that they sum to the target population in the 1,004 districts. The following adjustments were made within the 12 region-by-urbanicity categories.

<b>Exhibit E6</b>	
<b>The Sum of the Weights from the Sampled Schools</b>	
<b>School Level</b>	<b>Sum of Weights</b>
Unique Schools	30,022
Elementary Schools	20,427
Middle Schools	8,159
High Schools	5,678

Let  $w_{hj}^{adj}$  be the new weight for the  $j^{th}$  school in the  $h^{th}$  stratum after the adjustment. And let  $w_{hj}$  be the weight before adjustment. Let  $n_h$  be the population number of schools in the  $h^{th}$  stratum. Let  $m_h$  be the number of schools in the sample. Then the new adjusted weight for the  $j^{th}$  school in the  $h^{th}$  stratum is calculated as:

$$w_{hj}^{adj} = w_{hj} \times \frac{n_h}{\sum_{j=1}^{m_h} w_{hj}}$$

For unique schools we have

$$\sum_{h=1}^{12} n_h = 27,141, \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj} = 30,021.66, \quad \text{and} \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj}^{adj} = 27,141.$$

while for elementary schools we have

$$\sum_{h=1}^{12} n_h = 18,320, \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj} = 20,426.96, \quad \text{and} \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj}^{adj} = 18,320.$$

Similar adjustments were made for middle and high schools. The resulting adjusted weights are shown in Exhibit E7.

<b>Exhibit E7</b>	
<b>The Sums of Adjusted School Weights</b>	
<b>School Level</b>	<b>Sum of Adjusted Weights</b>
Unique Schools	27,141
Elementary Schools	18,320
Middle Schools	7,221
High Schools	5,263

### ***Adjustment of Weights for Unit Nonresponse***

A school became a nonresponder in three ways. First, a school's district may not have given permission to a school to be surveyed. This could have happened if a district did not respond at all, or if a district responded but still did not give permission. Second, a school's district may have given permission, but the school simply did not respond. Third, the school may have returned a questionnaire but answered so few questions that we categorized it as unit nonresponse. Regardless of how the school became a nonresponder, reweighting was carried out as noted in the Overview: for example, if half the elementary schools in the urban West district did not respond, then the weights of the responding elementary schools in the urban West doubled.

## Weights for S2-Schools

### *Conditional Weights for S2-Schools*

Three independent samples of schools were taken from our supplemental district sample: a sample of elementary schools, middle schools and high schools.

A sample of 1,272 elementary schools was selected from a sampling frame of 2,547 elementary schools. The weights for those 1,272 sampled elementary schools were obtained from the SAS program used to select the sample. The sum of the sampling weights for the 1,272 schools is 2,547. The weights for middle and high schools were obtained in the same way. The numbers of schools in the sampling frames, the numbers of schools selected and the sums of their conditional weights for elementary, middle and high schools are summarized in Exhibit E8.

<b>Exhibit E8</b>		
<b>Schools Selected from 626 Districts in Our Supplemental Sample</b>		
<b>School Level</b>	<b>Sums of Conditional Sampling Weights: Number of Schools in the Sampling Frame</b>	<b>Number of Schools Selected</b>
Elementary Schools	2,547	1,272
Middle Schools	1,071	884
High Schools	793	678

In addition to conducting analyses at the school level, we also wished to run analyses for all schools combined. This requires the calculation of the probability that a school was selected for *any* of the three samples, namely

$$\begin{aligned} \Pr(E \text{ or } M \text{ or } H) = & \Pr(E) + \Pr(M) + \Pr(H) \\ & - \Pr(E \text{ and } M) - \Pr(E \text{ and } H) - \Pr(M \text{ and } H) \\ & + \Pr(E \text{ and } M \text{ and } H). \end{aligned}$$

Because the three samples were independent, all joint probabilities were simply the products of the marginal probabilities. Thus  $\Pr(E \text{ and } M) = \Pr(E) \cdot \Pr(M)$ ,  $\Pr(E \text{ and } H) = \Pr(E) \cdot \Pr(H)$ ,  $\Pr(M \text{ and } H) = \Pr(M) \cdot \Pr(H)$ , and  $\Pr(E \text{ and } M \text{ and } H) = \Pr(E) \cdot \Pr(M) \cdot \Pr(H)$ .

Combining the sampling frames of elementary, middle and high schools in the 626 districts in our supplemental sample gave 3,940 unique schools. There were 2,556 unique schools selected in the elementary, middle or high school samples. These calculations were then summed over the 2,556 unique schools. The sum of the weights was 3,948.18, which was 8.18 too many because the 2,556 sampled schools were supposed to represent the 3,940 schools in the sampling frame. The weights were adjusted to sum to 3,940.

After adjusting the weights, it was ensured that no school had a weight of less than one. Out of the 2,556, there were 1,707 representing themselves only (had a sampling weight of one) and 849 purporting to represent the remaining 2,233 schools ( $3,940 - 1,707 = 2,233$ ). However, the sum of the weights for the 849 schools was 2,241.18 (8.18 too many), so weights were reduced by a factor of

2,233/2,241.18 for these schools. The new adjusted weights for the 2,556 schools summed to 3,940, as it should.

<b>Exhibit E9</b>		
<b>Schools Selected from 626 Districts in Our Supplemental Sample</b>		
<b>School Level</b>	<b>Sums of Conditional Sampling Weights: Number of Schools in the Sampling Frame</b>	<b>Number of Schools Selected</b>
Elementary Schools	2,547	1,272
Middle Schools	1,071	884
High Schools	793	678
All Schools Combined	3,940	2,566

***Unconditional Weights for S2-Schools***

Unconditional weights for S2-schools were calculated in a fashion similar to those for S1-schools. The probability of selection was

$$\Pr(s_{ij} \in S2) = \Pr(s_{ij} \in S2 | d_i \in S2) \times \Pr(d_i \in S2),$$

where  $d_i$  is district  $i$ , and  $s_{ij}$  is school  $j$  in district  $i$ .

The sum of the weight variable for S2-districts is 12,314 because the 626 S2-districts represent the 12,314 S2-districts.

The conditional probability of selection,  $\Pr(s_{ij} \in S2 | d_i \in S2)$  is the inverse of the adjusted weight for the All School Combined weights. For the elementary, middle and high school samples, this is simply the probability.

Exhibit E10 shows the population number of schools in the 12,314 districts. The sample of S2-schools is intended to represent the population of schools in the 12,314 districts.

<b>Exhibit E10</b>	
<b>Numbers of Schools in the 12,314 Districts</b>	
<b>School Level</b>	<b>Number of Schools</b>
Unique	55,806
Elementary	34,360
Middle	17,450
High	13,821

Exhibit E11 shows that the sum of the weights from the sampled schools was slightly lower than the number of schools in the 12,314 districts. School weights were adjusted in order that they sum to the

target population of schools in the 12,314 districts. These adjustments were calculated within the 12 region-by-urbanicity categories.

<b>Exhibit E11</b>	
<b>The Sum of the Weights from the Sampled Schools</b>	
<b>School Level</b>	<b>Number of Schools</b>
Unique	51,263
Elementary	31,130
Middle	16,047
High	12,744

Using a similar notation as in the adjustment of S1-schools, we let  $w_{hj}^{adj}$  be the new weight for the  $j^{th}$  school in the  $h^{th}$  stratum after the adjustment, and let  $w_{hj}$  be the weight before adjustment. Let  $n_h$  be the population number of schools in the  $h$ th stratum. Let  $m_h$  be the number of schools in the sample. Then the new adjusted weight for the  $j^{th}$  school in the  $h^{th}$  stratum is calculated as:

$$w_{hj}^{adj} = w_{hj} \times \frac{n_h}{\sum_{j=1}^{m_h} w_{hj}}$$

For unique schools we have

$$\sum_{h=1}^{12} n_h = 55,806, \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj} = 51,263.05, \quad \text{and} \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj}^{adj} = 55,806$$

and for elementary schools we have

$$\sum_{h=1}^{12} n_h = 34,360, \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj} = 31,129.75, \quad \text{and} \quad \sum_{h=1}^{12} \sum_{j=1}^{m_h} w_{hj}^{adj} = 34,360.$$

Similar adjustments were made for middle and high schools. The adjusted weights are shown in Exhibit E12.

<b>Exhibit E12</b>	
<b>Sums of Adjusted School Weights</b>	
<b>School Level</b>	<b>Number of Schools</b>
Unique	55,806
Elementary	34,360
Middle	17,450
High	13,821

## Appendix F. The District Nonresponse Study

The redesigned district sample described in Appendix B yielded a lower response rate than expected. To ensure that the resulting sample had no inherent bias, and to guarantee an acceptable level of precision, it was necessary for Abt to design a nonresponse study. This appendix describes the methodology we employed to design and conduct the study.

### Response Rates and the Potential for Bias

As noted in Appendix B, at the close of the district data collection effort we had obtained survey data from 298 districts.<sup>5</sup> That corresponds to either a 31 percent response rate or an 18 percent response rate, depending on what is used as the denominator. The two candidates for denominator are 1,630 and 959. The larger number is the number of districts selected in either the design phase or the implementation phase of the SLIIDEA contract. The smaller number is the number of districts that were ever actively pursued in the implementation phase. These are the only districts that ever received surveys in the mail.

In the design phase, the design contractor (AIR) had selected a sample of 1,004 districts. At the time that Abt and Westat began work on the implementation phase of the contract, AIR turned over a list of 324 districts that they had recruited to participate in the study. Given the expected difficulty of recruiting the remaining districts that had already indicated their unwillingness to participate, Abt and Westat made a decision to

- keep the 324 districts that had agreed to participate in the sample;
- attempt to recruit nine large districts from the original 1,004 (in order to fill some region-by-urbanicity-by-size cells);
- abandon efforts at recruiting the remaining 671 districts that AIR had selected but were unable to recruit; and
- select a supplemental sample of 626 new districts.

The new sample consisted of the 626 supplemental districts plus the 333 districts from AIR's original sample (324 + 9) for a total of 959 districts. These 959 districts were actively pursued in the implementation phase. All of these districts received surveys in the mail and most were contacted by phone on multiple occasions. From these 959 districts, we obtained 298 completed surveys, yielding the 31 percent response rate. The total number of districts that were ever selected was  $959 + 671 = 1,630$ . These 1,630 districts included 671 that were never mailed a survey. However, these 671 had indicated an unwillingness to participate. Including these districts in the denominator (i.e., dividing 298 by 1,630) results in the 18 percent response rate.

Using either candidate as the denominator, the response rates are very low. The low response rates present three major problems. The first is that we have fewer districts than we had originally planned on. This means that the precision of our estimates is not as good as it would have been with a better response rate. This is not a major problem for this study because a sample size of 298 is sufficiently

---

<sup>5</sup> Actually, instruments were received from 301 districts. But three of the districts had left such a large proportion of the survey incomplete that their data were deemed to be essentially useless, and they were classified as unit nonresponders.

large to give reasonably precise estimates, and the 298 districts are reasonably distributed across region, urbanicity and size categories. The second problem is that when districts do not participate, it often means that we can't survey the schools within the nonparticipating districts. Therefore the problem compounds for the school survey data. Rather than address that problem further in this document, we will address it in separate analyses. The third problem is that there is a large potential for bias. Bias occurs if the districts that refused to participate would have given systematically different responses to the survey (had they responded to it) than the districts that did complete and return the surveys. Poor response rates do not guarantee a biased sample. The decision to not participate could be completely unrelated to the answers to the survey questions. However, good response rates can make the effects of bias negligible even if bias exists. In this study, the response rates are such that we cannot ignore the potential for bias. It would be unreasonable to simply assume that nonparticipation is unrelated to the way in which districts would have answered the survey questions. We therefore conducted a nonresponse study to check for evidence of nonresponse bias.

A district could be classified as a nonresponder for three reasons:

1. During the recruitment phase, it could refuse to participate.<sup>6</sup>
2. The district could indicate a willingness to participate but then never fill out and return the survey.
3. The district could participate and mail back the survey, but leave so many of the survey items blank that the survey could be judged to be incomplete and the district could be classified as a nonresponder. Three districts were classified as nonresponders for this reason.

The plan for the district nonresponse study was to select a sample of nonresponding districts, then go back to them with a small subset of items from the original district questionnaire. We made an intensive effort to get them to respond to the small subset of items using telephone calls, faxes, etc. Most of the nonresponse survey items could be answered over the phone. A few of the items needed to be faxed to and from the districts.

The timing of the district nonresponse study presented a challenge. To select a sample of nonresponding districts, you have to know who the nonresponding districts are. If we had waited until the close of regular district data collection to assemble a list of nonresponding districts, little time would have been left during the school year to complete the survey. Instead, we used the large pool of nonresponding districts that were identified earlier in the year. Those were the 671 districts that were originally selected by AIR but that refused to participate and were not pursued by Westat and Abt. We took a sample of 133 districts from the 671 and obtained responses from 107 of them (a response rate of 80 percent). One conclusion that we immediately drew was that it is not impossible to get good response rates from districts when you have a very small, very low burden instrument that is administered mostly over the phone.

Making inferences about nonresponse bias from these 107 districts is dependent on *an important, but ultimately untested assumption*. That assumption is that the responses of districts that were originally

---

<sup>6</sup> During the implementation phase, districts were mailed questionnaires at the same time as the attempted recruitment, so they had a questionnaire whether they agreed to participate or not. If they did not agree to participate, however, we could not go to their schools with the school survey.

selected by AIR and refused to participate (nonresponse reason number 1, above) are representative of responses of

- districts from the AIR sample that agreed to participate but that did not return surveys (reason number 2);
- districts from the supplemental sample that agreed to participate but that did not return surveys (reasons number 2 and number 3); and
- districts from the supplemental sample that did not agree to participate (reason number 1).

Counting districts from both the design phase and the implementation phase, there were a total of  $1,630 - 298 = 1,332$  districts that either could not be recruited, or were recruited but never returned a survey.

## Districts That Responded to the Nonparticipation Survey

For the remainder of this document, the 107 districts that responded to the nonparticipation survey will be referred to as “NonPar” districts. The 298 that returned surveys in the core survey will be called the “Core” districts.

AIR’s sample of 1,004 districts included all districts in the country that had more than 40,500 students. Their sampling strategy oversampled for large districts in general. Therefore, we expected, a priori, that a subsample of the AIR sample of districts would likely be composed of districts that are larger, on average, than either the districts in the supplemental sample or than the average of all districts in the population. This expectation turned out to be true. Data from NonPar and Core districts were compared on several *frame variables*.<sup>7</sup> In summary, the results show that Core and NonPar districts were similar with respect to the following:

- Region (4 census regions)
- Urbanicity (urban, suburban, rural)
- Stratum (12 region-by-urbanicity strata)
- Orshansky Poverty Index (obtained from QED)
- Type of agency (types 1, 2, and 3 listed in Exhibit B1)
- Percent IEP (percent of students with a written Individual Education Program).

Core and NonPar districts were different with respect to the following:

- Size (of student population)—the NonPar districts were, on average, larger
- Number of schools—the NonPar districts, on average, had a greater number of schools.

In itself, the fact that the NonPar districts were larger, on average, than the Core districts, does not present major problems. Distributional differences in size, region, urbanicity or other variables that are known for all districts in the population can be controlled for. The important question is whether the estimates we obtain from the Core district data (e.g., means and percents) would have been substantively different if they had been calculated from responses of all selected districts rather than calculated from only the 298 responding districts. This is the question of bias.

---

<sup>7</sup> Frame variables are variables that are known for all districts in the population (e.g., region, urbanicity, size).

## Estimation of the Bias in Survey Responses

The 107 NonPar districts were administered a small subset of the full set of survey items that were administered to the Core districts. The survey questions in the district nonparticipation survey are shown in Appendix H.

An overview of the method for estimating bias in survey responses is as follows: First, for a particular survey item, calculate the mean and the standard error using only the data from the 298 Core districts. This calculation uses weighted data such that the 298 districts will represent all 13,318 districts in the target population. In most cases, the *mean* is a proportion and will be interpreted as “the proportion of districts in the target population that ... [offer financial assistance to schools, or regularly evaluate parent satisfaction, etc.]” Estimates calculated in this fashion are the unadjusted estimates that we would have used if we had never conducted a nonresponse study.

The second step is to calculate an estimate of the population mean using data from both the Core and the NonPar districts. This estimate depends on two assumptions: that there are two types of districts in the target population and that the two types may have systematically different answers to survey questions. The two types are those that would willingly respond to the survey and those that would not. In this estimate, the Core districts represent the population of districts that are willing to respond, the NonPar districts represent the population of districts that are unwilling to respond, and a weighted average of the two represents the entire target population of districts.

The third step is to calculate the difference between the estimates from the first and second steps. This difference is the estimated bias. The fourth step is to divide the bias by the standard error of the estimate from step one. If the bias is small relative to the standard error, then it is not substantial. One way to think of it is, if the bias is about 10 percent or less of the standard error, then it is small enough to ignore. The logic is that the bias is well within the 95 percent confidence interval and hence the estimate is still in the range you would have expected anyway.

### **Step 1: Estimating the Mean and Standard Error from the 298 Core Districts**

The calculation of the initial sampling weights for all 959 districts that received surveys is described in Appendix C. If all 959 districts had responded to the survey, the original sampling weights would have been used, and the weighted data would represent the target population of 13,318 districts. That is, the weights of all 959 districts sum to 13,318. But data were received from only 298 districts. The original weights for the 298 districts sum to a number that is far less than 13,318. Therefore, in order for those 298 districts to represent the entire target population, their weights must be adjusted. This adjustment rests on the assumption that nonresponding districts are no different than responding districts in the way that they would answer the questions on the survey. This adjustment, or reweighting for nonresponse, was done within 30 weighting classes representing categories of region, urbanicity and size. Within each weighting class, the sum of the weights is equal to the target population number of districts in the weighting class. The sum of the weights over all 30 classes is equal to the total number of districts in the target population (13,318).

The rationale for the choice of the 30 weighting classes is as follows: Originally, selection of the sample of districts was taken from within 12 region-by-urbanicity strata. Thus, reweighting within those strata would appear to be a reasonable approach. However, exploratory analyses indicated that district size is an important factor related to survey responses. And our analyses have shown that the Core districts are smaller, on average, than the target population of districts. We therefore wanted to

control for differences in the size distributions of Core districts and the target population when assessing nonresponse bias. Reweighting within size categories is a method to achieve that statistical control. We sought to create enough size categories to allow for reasonably precise balance of size between the weighted distributions of Core data and the combined Core+NonPar data. Furthermore, we wanted to use the same weighting classes in both Steps 1 and 2. The weighting classes were made from four region categories, three urbanicity categories and five size categories. There were not sufficient numbers of Core and NonPar districts in each of the possible  $4 \times 3 \times 5 = 60$  categories to have weighting classes for all 60 categories. Therefore, some collapsing of categories was necessary. The 30 weighting classes are shown in both Exhibits F1 and F2. The exhibits show that most of the collapsing of categories was done by combining urban and suburban districts within region and size groups. There were several classes that combined rural districts within a size category over two regions.

Exhibit F1 summarizes the Step 1 reweighting for nonresponse within the 30 weighting classes. For example, row 1 of the exhibit shows that the sum of the original weights within the weighting class labeled “ $\leq 1,500$  MW Sub” is 225.72. The weights of each of the nine districts within this weighting class were inflated by a factor of 3.9518. The sum of the new weights in that class is now 892 and is equal to the population number of districts in the class. The same process is repeated for districts in each weighting class. The sum of the resulting new weights is equal to the population number of districts, 13,318.

Using the new weights, we now show the mean and the standard error for an example survey item. For the example item, G1a, the mean corresponds to the proportion of districts that provide inservice training for administrators. The mean and the standard error were calculated using SUDAAN software. These estimates take into account the sampling strata and sampling weights.

Results for 298 Core districts on item G1a:

- *Estimate 1*: “Yes” response for 62.9 percent of districts
- Standard error = 4.8

**Exhibit F1**

**Step 1 Inflation of Weights Within Thirty Weighting Classes**

Weighting Class				N Core	Sum of Orig Wts	Correction Factor	Sum of New Wts	Pop N
1	<=1500	MW	Sub	9	225.72	3.9518	892	892
2	<=1500	NE	Sub	16	408.24	2.0405	833	833
3	<=1500	SO	Urb Sub	15	196.97	2.0054	395	395
4	<=1500	WE	Urb Sub	13	158.76	2.9731	472	472
5	<=1500	MW NE	Rur	19	817.48	3.4924	2,855	2,855
6	<=1500	SO	Rur	16	363.26	3.0749	1,117	1,117
7	<=1500	WE	Rur	4	103.01	8.4067	866	866
8	1501-3500	MW	Urb Sub	21	202.45	3.3983	688	688
9	1501-3500	NE	Urb Sub	11	125.49	5.4746	687	687
10	1501-3500	SO	Urb Sub	2	8.68	21.8861	190	190
11	1501-3500	WE	Sub	3	16.45	13.8563	228	228
12	1501-3500	MW	Rur	6	16.55	27.3187	452	452
13	1501-3500	NE	Rur	1	64.66	2.4125	156	156
14	1501-3500	SO	Rur	14	163.06	3.2259	526	526
15	1501-3500	WE	Rur	3	6.63	27.2937	181	181
16	3501-8000	MW	Urb Sub	20	141.87	2.5235	358	358
17	3501-8000	NE	Urb Sub	12	71.26	5.9220	422	422
18	3501-8000	SO	Urb Sub	19	79.17	2.9682	235	235
19	3501-8000	WE	Urb Sub	9	39.98	5.3025	212	212
20	3501-8000	MW NE	Rur	8	22.59	5.6665	128	128
21	3501-8000	SO WE	Rur	15	74.16	5.1914	385	385
22	8001-20000	MW	Urb Sub	10	44.20	3.5976	159	159
23	8001-20000	NE	Urb Sub	1	2.45	36.6667	90	90
24	8001-20000	SO	Urb Sub	16	59.78	3.0613	183	183
25	8001-20000	WE	Urb Sub	9	30.38	6.6491	202	202
26	8001-20000	SO WE	Rur	5	30.75	3.2192	99	99
27	>20000	MW	Urb Sub	5	12.92	2.8631	37	37
28	>20000	NE	Urb	3	7.36	2.1728	16	16
29	>20000	SO	Urb Sub	11	36.82	4.1830	154	154
30	>20000	WE	Urb Sub	2	6.52	15.3285	100	100
<b>Total</b>				<b>298</b>	<b>3,537.63</b>		<b>13,318</b>	<b>13,318</b>

Thirty weighting classes were used. Size: <=1500 means up to 1,500 students, 1501–2500 means 1,501 to 2,500 students, etc. Region: MW=Midwest, NE=Northeast, SO=South, WE=West. Urbanicity: Urb=Urban, Sub=Suburban, Rur=Rural. N Core is the number of Core districts in the weighting class. Sum of Orig Wts is the sum of the original weights of the core districts within a weighting class. Correction Factor is the amount that each district’s original weight was inflated, within a weighting class, to achieve the new weights. Sum of New Wts is the sum of the new weights, within a weighting class. Notice that the sum of the new weights is equal to the Pop. N (number of districts within a weighting class in the target population).

**Step 2: Estimating the Combined Mean Using Data from Both the 298 Core Districts and the 107 NonPar Districts**

We now estimate the mean using data from all 405 districts (298 Core districts and 107 NonPar districts). The weights for the 298 Core districts are their original weights (see Exhibit F2, column labeled Sum of Orig Wts). These districts represent all of the districts in the population that are the type that would participate. The weights of 107 NonPar districts are inflated to represent all of the

nonresponding districts. That is, they are weighted to represent all of the districts in the population that are the type that would not participate (see Exhibit F2, column labeled Sum of NonPar Wts). Together, these samples of two types of districts (those that are willing to participate and those that are not willing to participate) represent all districts in the population.

Using SUDAAN, the weighted estimate of the mean from all 405 districts on item G1a was:

- *Estimate 2*: “Yes” response for 71.0 percent of districts
- Standard error = 6.3

<b>Exhibit F2</b>								
<b>Step 2 Weights for Combined Sample of Core and NonPar Districts</b>								
<b>Weighting Class</b>				<b>N Core</b>	<b>Sum of Orig Wts</b>	<b>N NonPar</b>	<b>Sum of NonPar Wts</b>	<b>Pop N</b>
1	<=1500	MW	Sub	9	225.72	2	666.28	892
2	<=1500	NE	Sub	16	408.24	1	424.76	833
3	<=1500	SO	Urb	15	196.97	1	198.03	395
4	<=1500	WE	Urb	13	158.76	1	313.24	472
5	<=1500	MW NE	Rur	19	817.48	5	2,037.52	2855
6	<=1500	SO	Rur	16	363.26	1	753.74	1117
7	<=1500	WE	Rur	4	103.01	1	762.99	866
8	1501-3500	MW	Urb	21	202.45	4	485.55	688
9	1501-3500	NE	Urb	11	125.49	5	561.51	687
10	1501-3500	SO	Urb	2	8.68	2	181.32	190
11	1501-3500	WE	Sub	3	16.45	1	211.55	228
12	1501-3500	MW	Rur	6	16.55	4	435.45	452
13	1501-3500	NE	Rur	1	64.66	2	91.34	156
14	1501-3500	SO	Rur	14	163.06	4	362.94	526
15	1501-3500	WE	Rur	3	6.63	2	174.37	181
16	3501-8000	MW	Urb	20	141.87	5	216.13	358
17	3501-8000	NE	Urb	12	71.26	6	350.74	422
18	3501-8000	SO	Urb	19	79.17	5	155.83	235
19	3501-8000	WE	Urb	9	39.98	4	172.02	212
20	3501-8000	MW NE	Rur	8	22.59	1	105.41	128
21	3501-8000	SO WE	Rur	15	74.16	4	310.84	385
22	8001-20000	MW	Urb	10	44.2	4	114.8	159
23	8001-20000	NE	Urb	1	2.45	2	87.55	90
24	8001-20000	SO	Urb	16	59.78	3	123.22	183
25	8001-20000	WE	Urb	9	30.38	8	171.62	202
26	8001-20000	SO WE	Rur	5	30.75	2	68.25	99
27	>20000	MW	Urb	5	12.92	1	24.08	37
28	>20000	NE	Urb	3	7.36	2	8.64	16
29	>20000	SO	Urb	11	36.82	16	117.18	154
30	>20000	WE	Urb	2	6.52	8	93.48	100
<b>Total</b>				<b>298</b>	<b>3,537.63</b>	<b>107</b>	<b>9,780.37</b>	<b>13,318</b>

Thirty weighting classes were used. Size: <=1500 means up to 1,500 students, 1501–2500 means 1,501 to 2,500 students, etc. Region: MW=Midwest, NE=Northeast, SO=South, WE=West. Urbanicity: Urb=Urban, Sub=Suburban, Rur=Rural. N Core is the number of Core districts in the weighting class. Sum of Orig Wts is the sum of the original weights of the core districts within a weighting class. N NonPar is the number of NonPar districts within each weighting class. Sum of NonPar Wts is the sum of the inflated weights for NonPar districts, within each weighting class. Notice that the sum of the Orig Wts and the NonPar Wts within a weighting class is equal to the population number of districts within a weighting class.

### Step 3: Estimating the Bias

$$\begin{aligned}\text{Bias} &= \text{Estimate 1} - \text{Estimate 2} \\ &= 62.9\% - 71.0\% \\ &= -8.1\%\end{aligned}$$

### Step 4: Calculating the Ratio of the Bias to the Standard Error

$$\begin{aligned}\text{Ratio} &= \text{Bias} / \text{S.E. Core} \\ &= -8.1 / 4.8 \\ &= -1.7\end{aligned}$$

The results shown in Step 3 suggest that if we were to estimate the mean from the Core data only, we would be underestimating the population percentage of districts that provide inservice training for administrators by about 8 percentage points. The Step 4 results for the current example indicate a large negative bias (more than one and a half times the size of the standard error, much bigger than 10 percent).

## Results: Estimated Bias for Each of the Nonresponse Survey Items

Recall that we are only able to estimate bias for the items that were on both the Nonresponse survey and the Core survey instruments (the items are shown in Appendix H). The bias estimates for each of these items are summarized in Exhibits F3 through F8. Many of the bias estimates are quite large, ranging between one to six times the size of the standard errors of the estimates from the Core districts. Some summary statements about the results in Exhibits F3 through F8 follow.

The Core data underestimate the percent of districts that

- Provide technical or financial assistance to schools based on test scores.
  - Based on test scores of the general population of students.
  - Based on test score of students with disabilities.
- Conduct regular evaluations of parent satisfaction.
- Provide support to schools that are developing school-wide approaches to discipline
  - Each of eight types of support.
- In the prior year, have had one or more cases involving mediation, impartial due process, state administrative review or state/federal judicial review of hearing decision.
- Have more than five students per 1,000 who have been placed in interim alternative settings.
- Report the scores of students with disabilities on district-wide assessments either separately from the general population of students, or both aggregated and separately.

Regarding the use of content standards, the Core data underestimate the percent of districts that have no standards or have not made decisions on the use of content standards. At the same time, they tend to overestimate the percent of districts that have the same standards or separate, modified or enhanced standards. The patterns are less clear regarding the use of performance standards, and the biases tend

to be of a size that is modest relative to the other bias estimates. Most of the overestimates are less than the size of one standard error.

The directions of the biases are less consistent for the remaining items. Use of Core data only results in small overestimates of the percent of districts that

- Provide technical assistance to schools based on dropout rates.
  - Based on dropout rates of the general population of students.
  - Based on dropout rates of students with disabilities.

But underestimate the percent of districts that

- Provide financial assistance to schools based on dropout rates.
  - Based on dropout rates of the general population of students.
  - Based on dropout rates of students with disabilities.

The recalibration to reduce the nonresponse bias is presented in Appendix G.

**Exhibit F3**

**District Report on Use of Content Standards Estimated Nonresponse Bias (Items B1a-f)**

Question	Response	Core_Y	Core_Se	Cmbin_Y	Cmbin_Se	Bias	Ratio
<b>Mathematics</b>							
	NoStds	3.5	1.6	7.3	5.6	-3.8	-2.4
	SameStds	47.7	5.1	45.3	6.5	2.4	0.5
	SepModEn	44.7	5.2	37.2	6.2	7.6	1.5
	NoDecisn	4.1	1.5	10.3	5.5	-6.2	-4.2
	SampSize	271.0	.	376.0	.	.	.
<b>Reading</b>							
	NoStds	3.1	1.5	7.4	5.6	-4.2	-2.8
	SameStds	47.4	5.1	44.2	6.5	3.2	0.6
	SepModEn	45.4	5.2	38.2	6.3	7.3	1.4
	NoDecisn	4.1	1.5	10.3	5.5	-6.2	-4.2
	SampSize	271.0	.	376.0	.	.	.
<b>Science</b>							
	NoStds	6.9	2.4	8.2	5.6	-1.3	-0.5
	SameStds	46.1	5.1	42.4	6.3	3.7	0.7
	SepModEn	42.8	5.2	39.1	6.4	3.7	0.7
	NoDecisn	4.2	1.5	10.3	5.5	-6.1	-4.1
	SampSize	270.0	.	375.0	.	.	.
<b>Social Studies</b>							
	NoStds	7.3	2.5	8.3	5.5	-0.9	-0.4
	SameStds	45.5	5.1	42.3	6.3	3.2	0.6
	SepModEn	43.0	5.2	39.1	6.4	3.9	0.7
	NoDecisn	4.1	1.5	10.3	5.5	-6.1	-4.2
	SampSize	270.0	.	375.0	.	.	.
<b>Writing</b>							
	NoStds	5.8	2.3	8.9	5.6	-3.2	-1.4
	SameStds	47.6	5.1	41.6	6.3	5.9	1.2
	SepModEn	42.1	5.1	39.1	6.4	3.0	0.6
	NoDecisn	4.5	1.5	10.3	5.5	-5.8	-3.8
	SampSize	270.0	.	375.0	.	.	.
<b>Foreign Language</b>							
	NoStds	23.6	4.1	20.5	6.1	3.2	0.8
	SameStds	42.1	5.1	39.5	6.2	2.5	0.5
	SepModEn	27.2	4.9	25.3	5.4	1.8	0.4
	NoDecisn	7.1	2.1	14.7	6.1	-7.5	-3.6
	SampSize	268.0	.	370.0	.	.	.

**Response:** **NoStds** = No content standards for any students. **SameStds** = Same standards for all students. **SepModEn** = Separate, modified or enhanced standards for students with disabilities. **NoDecisn** = District has not made a decision on the use of content standards.

**SampSize** = The number of Core districts providing a valid response to the question (out of a possible 298 districts). And number of combined Core&NonPar districts providing a valid response to the question (out of a possible 405 districts).

**Core\_Y** = Weighted percent of Core districts responding "Yes" to the category shown in the row (Step 1 estimates). **Core\_se** = standard error. **Cmbin\_Y** = Weighted percent "Yes" responses of combined Core and NonPar districts (Step 2 estimates). **Cmbin\_se** = standard error. **Bias** = Core\_Y - Cmbin\_Yes. **Ratio** = Bias/Core\_se

**Exhibit F4****District Report on Use of Performance Standards Estimated Nonresponse Bias (Items B2a-f)**

Question	Response	Core_Y	Core_Se	Cmbin_Y	Cmbin_Se	Bias	Ratio
<b>Mathematics</b>							
	NoStds	6.9	2.3	10.1	5.8	-3.2	-1.4
	SameStds	39.5	4.9	39.7	6.0	-0.1	-0.0
	SepModEn	46.8	5.2	44.4	6.3	2.4	0.5
	NoDecisn	6.8	2.4	5.8	1.8	0.9	0.4
	SampSize	267.0	.	371.0	.	.	.
<b>Reading</b>							
	NoStds	6.4	2.3	10.0	5.8	-3.6	-1.6
	SameStds	39.5	4.9	39.6	6.0	-0.1	-0.0
	SepModEn	47.3	5.1	44.5	6.3	2.8	0.5
	NoDecisn	6.8	2.4	5.8	1.8	0.9	0.4
	SampSize	267.0	.	371.0	.	.	.
<b>Science</b>							
	NoStds	9.9	2.6	10.9	5.8	-1.0	-0.4
	SameStds	37.8	4.9	39.4	6.0	-1.6	-0.3
	SepModEn	45.1	5.2	41.0	6.2	4.1	0.8
	NoDecisn	7.2	2.4	8.7	3.1	-1.5	-0.6
	SampSize	266.0	.	369.0	.	.	.
<b>Social Studies</b>							
	NoStds	12.0	3.0	11.6	5.8	0.4	0.1
	SameStds	35.7	.7	38.8	6.0	-3.1	-0.7
	SepModEn	45.2	5.1	40.9	6.2	4.3	0.8
	NoDecisn	7.1	2.4	8.7	3.1	-1.6	-0.7
	SampSize	267.0	.	369.0	.	.	.
<b>Writing</b>							
	NoStds	7.5	2.4	11.2	5.8	-3.8	-1.6
	SameStds	39.0	4.9	38.6	6.0	0.4	0.1
	SepModEn	46.2	5.2	41.6	6.2	4.7	0.9
	NoDecisn	7.3	2.4	8.6	3.1	-1.3	-0.5
	SampSize	266.0	.	370.0	.	.	.
<b>Foreign Language</b>							
	NoStds	28.2	4.5	24.5	6.3	3.7	0.8
	SameStds	31.8	4.6	35.0	5.9	-3.2	-0.7
	SepModEn	29.6	4.9	28.5	5.6	1.1	0.2
	NoDecisn	10.4	2.9	12.0	4.0	-1.6	-0.6
	SampSize	267.0	.	365.0	.	.	.

**Exhibit F5****How Districts Report Scores of Students with Disabilities on District-wide Assessments  
Estimated Nonresponse Bias (Items B3, B4)**

Response	Core_Y	Core_Se	Cmbin_Y	Cmbin_Se	Bias	Ratio
<b>Question: How Report</b>						
Aggregated w/ students without disabilities	20.2	3.9	14.1	3.0	6.1	1.6
Separately	5.9	1.6	8.7	3.6	-2.7	-1.7
Both aggregated and separately	27.0	4.4	29.1	5.6	-2.1	-0.5
Not reported	1.6	0.8	0.6	0.2	1.1	1.4
No district-wides	45.3	5.1	47.6	6.9	-2.3	-0.5
Sample size	288.0	.	390.0	.	.	.

**Exhibit F6****Estimated Nonresponse Bias on Items B6a, B6b, B7a, B7b, H3a, H3b, H4a, H4b, J2, and K3 (See Appendix H)**

Question	Response	Core_Y	Core_Se	Cmbin_Y	Cmbin_Se	Bias	Ratio
Technical assistance/test scores/any	Yes	44.7	4.8	52.9	6.7	-8.3	-1.7
	SampSize	286.0	.	386.0	.	.	.
Financial assistance/test scores/any	Yes	25.8	3.9	33.5	5.9	-7.7	-2.0
	SampSize	284.0	.	382.0	.	.	.
Technical assistance/test scores/with focus on students with disabilities	Yes	32.3	4.3	44.9	6.7	-12.6	-2.9
	SampSize	279.0	.	378.0	.	.	.
Financial assistance/test scores/with focus on students with disabilities	Yes	19.2	3.6	28.6	5.7	-9.3	-2.6
	SampSize	277.0	.	372.0	.	.	.
Technical assistance/dropout/any	Yes	16.3	2.9	15.1	3.0	1.2	.04
	SampSize	273.0	.	372.0	.	.	.
Financial assistance / dropout/any	Yes	7.1	1.7	8.0	2.0	-0.9	-0.5
	SampSize	271.0	.	369.0	.	.	.
Technical assistance/dropout/with focus on students with disabilities	Yes	12.5	2.6	12.4	2.7	0.1	0.0
	SampSize	274.0	.	373.0	.	.	.
Financial assistance/dropout/with focus on students with disabilities	Yes	4.3	1.3	6.9	1.9	-2.5	-1.9
	SampSize	270.0	.	367.0	.	.	.
Regular evaluation of parent satisfaction	Yes	23.3	3.8	46.7	6.6	-23.4	-6.1
	SampSize	283.0	.	389.0	.	.	.
One or more Cases 0=0 cases 1=1 + cases	Yes	14.3	2.4	17.5	3.9	-3.2	-1.4
	SampSize	276.0	.	381.0	.	.	.

<b>Exhibit F7</b>							
<b>District Support for Schools That Are Developing School-wide Approaches to Discipline: Estimated Nonresponse Bias (Items G1a-h)</b>							
<b>Question</b>	<b>Response</b>	<b>Core_Y</b>	<b>Core_Se</b>	<b>Cmbin_Y</b>	<b>Cmbin_Se</b>	<b>Bias</b>	<b>Ratio</b>
Inservice training for administrators	Yes	62.9	4.8	71.0	6.3	-8.1	-1.7
	SampSize	284.0	.	390.0	.	.	.
Inservice training for general education teachers	Yes	64.7	5.2	81.6	3.3	-16.9	-3.3
	SampSize	282.0	.	385.0	.	.	.
Inservice training for special education teachers	Yes	75.2	4.5	87.3	2.7	-12.1	-2.7
	SampSize	281.0	.	385.0	.	.	.
Inservice training for other	Yes	59.2	5.3	62.0	6.6	-2.8	-0.5
	SampSize	282.0	.	387.0	.	.	.
Guidance or consultation from district staff	Yes	65.9	5.3	73.0	5.8	-7.1	-1.3
	SampSize	281.0	.	387.0	.	.	.
Funds for additional security	Yes	33.7	4.7	34.2	5.8	-0.4	-0.1
	SampSize	278.0	.	383.0	.	.	.
Program for developing partnerships	Yes	43.7	5.0	45.7	6.5	-2.0	-0.4
	SampSize	279.0	.	385.0	.	.	.
Help developing partnerships	Yes	50.7	5.2	53.4	6.7	-2.7	-0.5
	SampSize	281.0	.	387.0	.	.	.

<b>Exhibit F8</b>							
<b>Placement of Students with Disabilities in Interim Alternative Settings (Item G4)</b>							
<b>Estimated Nonresponse Bias</b>							
<b>Question</b>	<b>Response</b>	<b>Core_Y</b>	<b>Core_Se</b>	<b>Cmbin_Y</b>	<b>Cmbin_Se</b>	<b>Bias</b>	<b>Ratio</b>
<b>Number in Interim Setting</b>	0 interim setting	43.7	5.4	38.7	6.3	5.0	0.9
	>0 – 0.5 per 1000	6.7	1.8	4.6	1.2	2.1	1.2
	>0.5 – 2 per 1000	19.3	3.3	19.4	4.7	-0.1	-0.0
	>2 – 5 per 1000	15.9	4.0	16.7	6.1	-0.8	-0.2
	>5 per 1000	14.4	4.1	20.6	6.6	-6.2	-1.5
	SampSize	242.0	.	340.0	.	.	.
	Responses to question regarding the number of students with disabilities that were placed in interim alternative settings were converted to the number of students placed in interim alternative settings per 1,000 students in the district.						



## Appendix G: Recalibrating Weights to Reduce Nonresponse Bias

After the nonresponse surveys were designed and administered, it became apparent from the results that there were some systematic differences between the answers given by Core and NonPar districts to items on the nonresponse survey. The results indicated that the many of the estimates calculated from the 298 Core districts were biased relative to the combined estimates calculated from the Core and the NonPar districts (405 districts). For example, we found that the Core data underestimated the percent of districts that provide technical assistance to schools based on the test scores of students with disabilities (survey item B7a). Exhibit F6 showed that for Item B7a, the estimate from the 298 Core districts was 32.3 percent and the estimate from the combined sample of Core and NonPar districts was 44.9 percent. We refer to the former estimate as a *biased* estimate and the latter as *unbiased*, because the former ignores the finding that nonresponders give systematically different answers, and the latter includes the information from nonresponders.

In this section we explain a method for using the results of the nonparticipation survey to adjust the weights such that some of the biases associated with estimates obtained from the 298 Core districts will be reduced or eliminated.

In order to provide an accurate explanation of the method, we need to introduce some notation. Let  $\bar{Y}$  represent a biased estimate of the population mean. In the previous example of Item B7a,  $\bar{Y}$  was equal to 32.3%. Let  $\hat{X}$  represent the unbiased estimate of the population mean. In the example of item B7a this was 44.9%. Equation 1.1 (below) shows how the data from the 298 Core districts are used to arrive at the estimated population mean,  $\bar{Y}$ . The idea with recalibrating the weights is to adjust the weights used in Equation 1.1 so that the resulting mean will be equal to the unbiased estimate of the mean ( $\hat{X}$ ). More generally, if we have several variables for which we have unbiased estimates ( $\hat{X}_1, \hat{X}_2, \hat{X}_3, \dots$  etc.), we wish to recalibrate the weights such that results from Equation 1.1 will yield unbiased estimates for each of the variables. Recalibrating the weights such that they simultaneously result in unbiased estimates for each of several variables is an iterative process. The initial weights are first recalibrated such that  $\bar{Y}_1$  will equal  $\hat{X}_1$ . The resulting set of weights is next readjusted such that  $\bar{Y}_2$  will equal  $\hat{X}_2$ . We note, however, that after this second readjustment  $\bar{Y}_1$  will no longer be exactly equal  $\hat{X}_1$ . The next step is to move on to the third variable, then the fourth, etc. Then the whole process is repeated over and over again until the weights converge to a set of stable values that no longer change with each iteration.

The Core district data consist of the survey responses of 298 districts. We use the following equation to estimate the mean of a variable,  $y$ :

$$\bar{Y} = \frac{\sum_i w_i^0 I_{NM,i} y_i}{\sum_i w_i^0 I_{NM,i}}, \quad \text{[Eqn. 1.1]}$$

where,  $i = 1, 2, 3 \dots 298$ , and

$w_i^0$  is the sampling weight for the  $i^{\text{th}}$  district (the  $w_i^0$  are the weights calculated in Step 1 of Appendix C), and

$I_{NM,i}$  is an indicator for a nonmissing response on variable  $y$  for the  $i^{\text{th}}$  district, that is,

$$I_{NM,i} = \begin{cases} 1 & \text{if } y_i \text{ is nonmissing, and} \\ 0 & \text{if } y_i \text{ is missing, and} \end{cases}$$

$y_i$  is the response to a questionnaire item for the  $i^{\text{th}}$  district. For the remainder of this discussion, it will be convenient to assume that  $y_i$  is a binary response variable that takes the value 1, if a district answers “Yes” to a question, and takes the value 0 otherwise.

## Starting Out: The First Recalibration

Suppose we begin the recalibration using a variable,  $y$ , created from survey item B7a. This variable takes the value 1 if a district offers technical assistance to schools based on the test score results of students with disabilities, and takes the value zero otherwise. Here we will be starting with the initial sampling weights and recalibrating such that  $\overline{Y}_1$  will equal  $\overline{X}_1$ . To simplify the notation in this section, we will drop the “1”s from the subscripts (these subscripts indicate that we are referring to outcome variable number 1).

Let the recalibrated weights from the first iteration be denoted  $w_i^1$ . Note that with respect to the variable  $y$ , there are three types of districts: those that answered “Yes” ( $y = 1$  and  $I_{NM,i} = 1$ ), those that answered “No” ( $y = 0$  and  $I_{NM,i} = 1$ ), and those that did not respond to this item ( $y = .$  and  $I_{NM,i} = 0$ ).

For those districts that answered “Yes” ( $y = 1$  and  $I_{NM,i} = 1$ ),

$$w_i^1 = \frac{\overline{X}}{\widehat{Y}} w_i^0$$

Note that for these data, the means are proportions and in most cases the unbiased mean was larger than the biased mean ( $\overline{X} > \widehat{Y}$ ). Therefore, in such cases (i.e., when  $\overline{X} > \widehat{Y}$ ), the formula is such that districts that answered “Yes” will have their weights inflated.

For those districts that answered “No” ( $y = 0$  and  $I_{NM,i} = 1$ ),

$$w_i^1 = \frac{(1 - \overline{X})}{(1 - \widehat{Y})} w_i^0$$

For the typical case (i.e., when  $\overline{X} > \widehat{Y}$ ), the equation above means that the districts that answered “No” will have their weights reduced.

For those that did not respond to this item ( $y = .$  and  $I_{NM,i} = 0$ ).

$$w_i^1 = w_i^0$$

The weights of the districts that did not respond to this item were not adjusted at all.

Putting all three pieces into one expression results in the following:

$$w_i^1 = \frac{\overline{\hat{X}}}{\overline{\hat{Y}}} w_i^0 y_i I_{NM.i} + \frac{(1 - \overline{\hat{X}})}{(1 - \overline{\hat{Y}})} w_i^0 (1 - y_i) I_{NM.i} + w_i^0 (1 - I_{NM.i})$$

A nice feature of this recalibration method is that the weights still sum to the population total number of districts. That is,

$$\sum_i w_i^1 = \sum_i w_i^0 = 13,318.$$

Substituting  $w_i^1$  for  $w_i^0$  in Equation 1.1 gives

$$\overline{\hat{Y}}^* = \frac{\sum_i w_i^1 I_{NM.i} y_i}{\sum_i w_i^1 I_{NM.i}} = \overline{\hat{X}}$$

## Subsequent Recalibrations

Next we recalibrate the weights based on the estimates from a second outcome variable, say, survey item G1b. This variable takes the value 1 if a district offers inservice training for general education teachers and takes the values zero otherwise. For this second variable, we will be readjusting the weights such that  $\overline{\hat{Y}}_2$  will equal  $\overline{\hat{X}}_2$ .

The recalibration process is the same as was described for the first variable, except that the input weights are those that were recalibrated in the previous step ( $w_i^1$ ), and the output weights are a new set of recalibrated weights, denoted  $w_i^2$ . They are calculated as

$$w_i^2 = \frac{\overline{\hat{X}}}{\overline{\hat{Y}}} w_i^1 y_i I_{NM.i} + \frac{(1 - \overline{\hat{X}})}{(1 - \overline{\hat{Y}})} w_i^1 (1 - y_i) I_{NM.i} + w_i^1 (1 - I_{NM.i}), \text{ where, of course,}$$

$y_i$ ,  $\overline{\hat{Y}}$ ,  $\overline{\hat{X}}$ , and  $I_{NM.i}$  now correspond to the second variable (created from district survey item G1b).

A total of eight variables were used in the recalibration process. After recalibration on the eighth variable, the process was started over again on the first variable. The process was iterated until convergence.

## Recalibration Based on Multicategory Response Variables

Several of the variables that were used for the recalibrations were of the multicategory response type. For example, a variable concerning district use of content standards for students with disabilities was coded such that 1 = no standards for any student; 2 = same standards for all students; 3 = separate, modified or enhanced standards for students with disabilities; and 4 = no decision on standards. The recalibration logic is similar to that described above for binary variables. However, to show the formula used for multicategory response variables, we will need to introduce some new notation.

Let  $\overline{\hat{X}}^{=1}$  be an unbiased estimate of the proportion of districts that give response  $y = 1$ , and let  $\overline{\hat{X}}^{=2}$  be an unbiased estimate of the proportion of districts that give response  $y = 2$ , and so on. Let  $\overline{\hat{Y}}^{=1}$  be the biased estimate of the proportion of districts that give response  $y = 1$ , and let  $\overline{\hat{Y}}^{=2}$  be the biased estimate of the proportion of districts that give response  $y = 2$ , and so on. Let  $y_i^{=1}$ , indicate a response of “no standards” for the  $i^{\text{th}}$  districts, and so on.

Then, for an example of a four-category variable (and assuming it is the third-variable recalibration process), the weight is calculated as:

$$w_i^3 = \frac{\overline{\hat{X}}^{=1}}{\overline{\hat{Y}}^{=1}} w_i^2 y_i^{=1} I_{NM.i} + \frac{\overline{\hat{X}}^{=2}}{\overline{\hat{Y}}^{=2}} w_i^2 y_i^{=2} I_{NM.i} + \frac{\overline{\hat{X}}^{=3}}{\overline{\hat{Y}}^{=3}} w_i^2 y_i^{=3} I_{NM.i} + \frac{\overline{\hat{X}}^{=4}}{\overline{\hat{Y}}^{=4}} w_i^2 y_i^{=4} I_{NM.i} + w_i^2 (1 - I_{NM.i})$$

## Results of the Recalibration Process

In this section we show the results of recalibrating the weights based on eight of the 29 survey items that were common to both the Core and the nonparticipation surveys. The results show that for the eight items that served as the basis of recalibration, the biases were reduced to zero or near zero. The results also show that the biases for the remaining 21 items were generally reduced to less than around three percentage points above or below zero. There were, however, several items with slightly larger biases (the largest was 5.6 percentage points), and in some cases the bias increased a little (e.g., from half a percentage point below zero to two points above). Overall, the process appears to have reduced the biases considerably. In order for these results to be meaningful for the overall study, we must make the following assumption. That is, that the evidence that recalibrating the weights has reduced the biases among the smaller set of survey items that were common to both the Core and nonparticipation surveys, implies that nonresponse biases among the much larger set of items on the Core survey have likewise been reduced.

The results are summarized in Exhibits G1 and G2. Exhibit G1 shows the results at the first iteration. Exhibit G2 shows the results after 25 iterations. The column labeled “NAME” shows the variable name. The variables are listed in the same order that they were shown in Exhibits F3 through F8 of Appendix B. The first variable, corresponding to item B1a, is shown as four rows (CoMaNoS, CoMaSam, CoMaSep, CoMaNoD) because it is a multicategory response variable with four possible responses. The labels for the four rows are shorthand for Content Standards Math, No Standard; Content/Math/Same Standards; Content/Math/Separate Standards; Content/Math/No Decision made. The column labeled “N” is the number of Core districts, out of a possible 298, that gave a valid response to the item. The column labeled “Core0” shows the estimated mean ( $\overline{\hat{Y}}$ ) using the initial sampling weights ( $w_i^0$ ). The column labeled Cmbin\_Y shows the unbiased mean ( $\overline{\hat{X}}$ ). The column labeled Bias0 shows the difference between  $\overline{\hat{Y}}$  and  $\overline{\hat{X}}$ , when the initial sampling weights are used.

The column labeled Bias1 shows the bias estimates after recalibrating on item B4. The rows corresponding to item B4 are labeled Agged, Sep, Both, NotRep, and NoDWid (rows 51–55). The row labels are shorthand for how districts report scores of students with disabilities: aggregated with other students, separately, both aggregated and separately, not reported, and district does not have district-wide assessments. Notice that the bias estimates for these rows are all zero in this column.

The column labeled Bias2 shows the bias estimates after recalibrating on item B7a (technical assistance based on test scores of students with disabilities, row 59). Notice that the bias estimates for these rows are all zero in this column. The bias estimates shown in the rows corresponding to item B4 (rows 51–55) are no longer equal to zero, but they are a lot closer to zero than they were in the column labeled Bias0.

Bias3 corresponds to item G1b (inservice training for general education teachers, row 68).  
Bias4 corresponds to item B1b (content standards for math, rows 3–6).  
Bias5 corresponds to item G4 (placement of students in interim alternative settings, rows 75–79).  
Bias6 corresponds to item J2 (regular evaluation of parent satisfaction, row 65).  
Bias7 corresponds to item G1d (inservice training for other staff, row 70).  
Bias8 corresponds to item B2e (performance standards in social studies, rows 39–42).

The rows corresponding to the calibration variables (B4, B7a, G1b, B1b, G4, J2, G1d, and B2e) are shaded in the exhibits.

The rows and columns have the same labels in Exhibits G1 and G2. Notice that after 25 iterations, the bias estimates are stable across columns Bias1–Bias8. The largest bias is 5.6 percentage points for item G1f (row 72).

Exhibit G3 is a graphical display of the amount of bias reduction. Line segments connect bias estimates before and after recalibration. The left-hand panel shows results for all of the variables (i.e., all the estimates in the column labeled Bias0 and all of the estimates in the column labeled Bias8 of Exhibits G1 and G2). All of the calibration variables have zero or near zero bias after recalibrating. In the right-hand panel of the exhibit, the calibration variables are omitted. This helps to illuminate the effect of recalibration on the variables that were not the basis of recalibration. This, in turn, may give some hint as to the expected effect of recalibration on the amount of bias on all of the items in the Core survey that were not also included in the nonresponse survey.

In conclusion, the method described in this document appears to have reduced biases associated with nonresponse among the items that were present on both the Core and the nonresponse surveys. There is no way of knowing for sure whether or not we have substantially reduced biases on items that were not on both surveys. However, for analyses of the Core data, we will use the weights derived by the methods described in this document and operate under the assumption that nonresponse bias has been reduced.

**Exhibit G1**

**Bias Estimates at Iteration 1**

Varno	Name	N	Core0	Cmbin_Y	Bias0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
1	_TYPE_	0	0.0000	.	.	.	.	.	.	.	.	.	.
2	_FREQ_	405	40500.00 00	.	.	.	.	.	.	.	.	.	.
3	CoMaNoS	271	3.4700	7.26	-3.78720	-3.51690	-4.20090	-3.66220	0.00000	0.08471	0.94101	1.63597	-0.59902
4	CoMaSam	271	47.7100	45.30	2.40833	1.15794	0.56930	-2.49520	0.00000	-1.40728	-4.69664	-5.16943	0.79731
5	CoMaSep	271	44.7400	37.16	7.57518	8.43694	10.23850	12.36750	0.00000	1.31670	3.56046	4.15171	1.87894
6	CoMaNoD	271	4.0800	10.28	-6.19630	-6.07798	-6.60690	-6.21000	0.00000	0.00586	0.19517	-0.61825	-2.07722
7	CoRdNoS	271	3.1400	7.35	-4.21357	-3.97360	-4.59060	-4.11240	-0.98072	-0.72475	0.33220	0.78210	-1.40156
8	CoRdSam	271	47.3600	44.20	3.16455	1.82491	1.15850	-2.00910	0.45030	-0.87344	-4.00491	-4.42085	1.59863
9	CoRdSep	271	45.4200	38.16	7.25532	8.23666	10.04900	12.34160	0.54041	1.60233	3.48754	4.26700	1.89015
10	CoRdNoD	271	4.0800	10.28	-6.19630	-6.07798	-6.60690	-6.21000	0.00000	0.00586	0.19517	-0.61825	-2.07722
11	CoScNoS	270	6.9000	8.19	-1.28997	-1.02114	-1.95450	-1.76850	1.89768	1.76148	2.11232	2.87968	0.04256
12	CoScSam	270	46.1100	42.38	3.73464	2.08988	1.63920	-1.40670	0.77710	-0.46211	-3.34809	-3.83068	2.63920
13	CoScSep	270	42.8100	39.12	3.69219	4.94475	6.81840	9.25470	-2.82239	-1.47899	0.91087	1.46042	-0.69922
14	CoScNoD	270	4.1700	10.31	-6.13686	-6.01349	-6.50300	-6.07960	0.14760	0.17962	0.32491	-0.50942	-1.98254
15	CoSsNoS	270	7.3400	8.25	-0.90881	-0.72059	-0.13570	0.60470	3.77897	4.11386	3.81136	3.89677	0.62541
16	CoSsSam	270	45.5400	42.33	3.20719	1.63288	-0.05750	-3.59220	-1.53352	-3.17752	-5.35753	-5.43435	1.39355
17	CoSsSep	270	42.9800	39.12	3.86070	5.12587	6.72980	9.11510	-2.26017	-1.00991	1.28294	2.09391	-0.02193
18	CoSsNoD	270	4.1400	10.29	-6.14907	-6.02816	-6.52660	-6.11760	0.02472	0.08358	0.27324	-0.54633	-1.98703
19	CoWrNoS	270	5.7700	8.92	-3.15072	-2.76426	-3.88780	-4.38050	-1.53271	-1.18813	-0.42816	0.29939	-2.12645
20	CoWrSam	270	47.5900	41.65	5.94241	4.43632	3.73720	0.70210	3.06828	1.72823	-1.49769	-1.95142	4.13106
21	CoWrSep	270	42.1400	39.13	3.00764	3.98869	6.41890	9.47690	-1.84786	-0.77637	1.55626	2.12455	-0.06269
22	CoWrNoD	270	4.5000	10.31	-5.80933	-5.67076	-6.27830	-5.80860	0.30228	0.22628	0.35958	-0.48253	-1.95192

**Exhibit G1**

**Bias Estimates at Iteration 1**

Varno	Name	N	Core0	Cmbin_Y	Bias0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
23	CoFLNoS	268	23.6100	20.46	3.15342	2.70938	1.78340	2.15470	6.29510	5.85767	5.30243	5.75116	3.15175
24	CoFLSam	268	42.0800	39.54	2.54147	1.45409	-0.42010	-2.81070	-3.26099	-4.52534	-6.21956	-6.77942	-1.60520
25	CoFLSep	268	27.1600	25.33	1.83296	3.41947	6.30870	8.68590	0.20033	1.92073	3.14703	4.19271	2.70655
26	CoFLNoD	268	7.1400	14.67	-7.52785	-7.58294	-7.67200	-8.02990	-3.23444	-3.25305	-2.22990	-3.16445	-4.25310
27	PrMaNoS	267	6.9000	10.06	-3.16033	-2.79428	-3.80420	-3.31880	0.31067	0.33847	1.04033	1.79213	-1.33594
28	PrMaSam	267	39.5400	39.67	-0.13438	-2.17350	-1.21150	-2.37330	-0.99528	-3.03626	-6.07090	-7.59639	1.95142
29	PrMaSep	267	46.8000	44.43	2.36606	3.81783	4.84200	5.77960	-3.70028	-2.13813	-0.14183	1.18342	-3.13261
30	PrMaNoD	267	6.7700	5.84	0.92865	1.14995	0.17370	-0.08740	4.38490	4.83592	5.17240	4.62084	2.51712
31	PrRdNoS	267	6.4000	10.04	-3.63946	-3.31091	-4.22300	-3.82550	-0.73202	-0.52947	0.41547	0.93971	-2.10668
32	PrRdSam	267	39.5400	39.61	-0.07438	-2.11350	-1.15150	-2.31330	-0.93528	-2.97626	-6.01090	-7.53639	2.01142
33	PrRdSep	267	47.3000	44.51	2.78520	4.27447	5.20080	6.22620	-2.71760	-1.33019	0.42304	1.97584	-2.42186
34	PrRdNoD	267	6.7700	5.84	0.92865	1.14995	0.17370	-0.08740	4.38490	4.83592	5.17240	4.62084	2.51712
35	PrScNoS	266	9.9300	10.91	-0.98464	-0.78876	-1.52150	-1.45210	1.82456	1.55766	1.91763	2.97642	-0.71578
36	PrScSam	266	37.8400	39.40	-1.56221	-3.51951	-2.83130	-3.84770	-2.42536	-4.17288	-6.97765	-8.76878	1.29416
37	PrScSep	266	45.0700	40.98	4.09248	5.60588	6.70120	8.05590	-1.13276	0.43672	2.39669	3.56253	-0.60812
38	PrScNoD	266	7.1644	8.71	-1.54560	-1.29760	-2.34840	-2.75600	1.73360	2.17850	2.66333	2.22980	0.02974
39	PrSsNoS	267	11.9959	11.55	0.44590	0.63720	1.00960	1.62620	4.49570	4.68300	4.02048	4.40260	0.00116
40	PrSsSam	267	35.7027	38.82	-3.11730	-5.07170	-5.19440	-6.69900	-5.46730	-7.59370	-9.29037	-10.69400	0.00388
41	PrSsSep	267	45.1961	40.93	4.26610	5.78610	6.60100	7.90620	-0.61880	0.85600	2.69847	4.13610	0.00409
42	PrSsNoD	267	7.1052	8.69	-1.58480	-1.34160	-2.40630	-2.82340	1.60040	2.06470	2.58143	2.16530	0.00087
43	PrWrNoS	266	7.4551	11.23	-3.77490	-3.33410	-4.47010	-4.37150	-1.33660	-1.15260	-0.34913	0.24720	-2.85546
44	PrWrSam	266	39.0345	38.60	0.43450	-1.60450	-0.58150	-1.68900	-0.33170	-2.36430	-5.37233	-6.91140	2.64218

**Exhibit G1**

**Bias Estimates at Iteration 1**

Varno	Name	N	Core0	Cmbin_Y	Bias0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
45	PrWrSep	266	46.2396	41.56	4.67960	6.09320	7.26470	8.47430	-0.28700	1.19330	3.11827	4.63860	0.31005
46	PrWrNoD	266	7.2707	8.61	-1.33930	-1.15460	-2.21320	-2.41390	1.95530	2.32360	2.60319	2.02570	-0.09677
47	PrFLNoS	267	28.2148	24.52	3.69480	3.41280	2.61200	2.01410	5.99960	5.70000	3.87568	4.96190	1.38503
48	PrFLSam	267	31.7841	34.99	-3.20590	-4.98620	-5.24790	-5.48670	-5.65190	-7.66540	-8.54893	-10.02920	-2.26352
49	PrFLSep	267	29.6240	28.51	1.11400	3.08170	4.63130	6.46300	-0.57890	1.50190	3.42366	4.43480	1.69881
50	PrFLNoD	267	10.3771	11.99	-1.61290	-1.51830	-2.00550	-3.00040	0.22120	0.45350	1.23960	0.62250	-0.83032
51	Agged	288	20.2254	14.13	6.09540	0.00000	0.47620	0.09100	-0.21000	-0.49340	-1.26847	-1.79070	-0.55305
52	Sep	288	5.9229	8.65	-2.72710	0.00000	-0.79830	0.52980	0.25980	0.20520	0.65402	0.43450	-0.01560
53	Both	288	26.9533	29.08	-2.12670	0.00000	0.85050	-0.58900	1.79290	1.65910	-3.11340	-2.36340	-2.21478
54	NotRep	288	1.6149	0.56	1.05490	0.00000	-0.09080	-0.13730	-0.14980	-0.12740	-0.24774	-0.22520	-0.23013
55	NoDWid	288	45.2835	47.58	-2.29650	0.00000	-0.43760	0.10550	-1.69290	-1.24350	3.97560	3.94490	3.01356
56	sepboth	288	32.8762	37.73	-4.85380	0.00000	0.05220	-0.05920	2.05260	1.86430	-2.45939	-1.92890	-2.23039
57	yb6a	286	44.6710	52.93	-8.25900	-8.32720	2.17380	5.31520	-0.09230	0.90300	-0.12962	-0.65310	0.09530
58	yb6b	284	25.7611	33.47	-7.70890	-8.07710	-1.56710	0.21240	-2.49650	-2.06360	0.72153	-0.11530	0.69729
59	yb7a	279	32.2722	44.88	-12.60780	-13.01470	0.00000	2.79090	-1.41150	-1.13680	0.10712	-0.88160	0.20857
60	yb7b	277	19.2304	28.58	-9.34960	-9.90970	-2.11650	0.00830	-3.61340	-2.76720	1.19041	0.71930	1.22928
61	yh3a	273	16.2828	15.05	1.23280	0.58610	2.78410	4.46600	3.55830	3.88670	2.63576	1.25720	2.35170
62	yh3b	271	7.1206	8.05	-0.92940	-0.74240	0.64200	1.58900	1.56630	1.33490	0.53279	-0.23290	0.85184
63	yh4a	274	12.4742	12.36	0.11420	-0.46930	1.49750	2.91020	2.43340	2.44540	2.16210	1.08130	2.07230
64	yh4b	270	4.3290	6.86	-2.53100	-2.79850	-1.39530	-0.98010	-0.83490	-1.09010	-0.99504	-1.38150	-0.69652
65	yj2	283	23.2959	46.73	-23.43410	-22.94920	-21.58570	-18.98580	-20.01290	-19.38880	0.00000	-1.31520	-2.38001
66	k3onepl	s 276	14.2981	17.51	-3.21190	-3.15040	-2.60210	-2.70600	-2.97870	-4.08120	-3.62343	-3.74380	-3.70655

**Exhibit G1**

**Bias Estimates at Iteration 1**

Varno	Name	N	Core0	Cmbin_Y	Bias0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
67	nG1A	284	62.8625	70.95	-8.08750	-8.69940	-6.24480	1.21220	1.75150	1.36410	2.20502	-3.39700	-1.19132
68	nG1B	282	64.7359	81.64	-16.90410	-16.14740	-13.73640	0.00000	0.46840	0.41330	3.08926	-1.20000	-2.23660
69	nG1C	281	75.1852	87.25	-12.06480	-11.41180	-9.65810	-0.11150	0.40600	0.01580	1.92091	-1.62260	-1.58948
70	nG1D	282	59.2439	62.00	-2.75610	-3.13330	-0.86480	8.66540	8.39250	7.81430	9.94319	0.00000	1.60172
71	nG1E	281	65.9008	73.02	-7.11920	-7.87660	-4.90140	2.12490	2.82440	3.21790	5.70931	2.00230	2.30468
72	nG1F	278	33.7117	34.16	-0.44830	-1.58470	-1.04800	1.67870	3.25930	3.36760	6.61890	4.76050	5.20803
73	nG1G	279	43.7070	45.71	-2.00300	-3.52160	-1.42340	2.54070	3.36480	2.30570	3.27112	-0.77960	-0.84785
74	nG1H	281	50.6528	53.39	-2.73720	-3.91770	-1.25420	4.02320	2.35160	1.54630	2.79338	-1.09660	-0.38190
75	Zero	242 4	3.6502	38.69	4.96020	4.72346	3.41142	0.24051	1.11585 7	0.00000	-0.83626	0.52662	0.99226
76	Gt0LEp5	242	6.7379	4.59	2.14789	1.86924	1.50488	1.74537	1.59409 1	0.00000	-0.10361	-0.50556	0.29517
77	Gtp5LE2	242 1	9.3141	19.44	-0.12591	-0.09232	0.84508	3.03798	3.73244 7	0.00000	-0.90004	-1.85071	-0.49285
78	Gt2LE5	242 1	5.8842	16.69	-0.80583	-0.60767	-2.04015	-2.70797	-2.36516	0.00000	-0.93601	-1.96199	-3.28352
79	Gt5	242 1	4.4136	20.59	-6.17636	-5.89271	-3.72123	-2.31588	-4.07722 1	0.00000	2.77593	3.79164	2.48894

“NAME” shows the variable name. “N” is the number of Core districts that gave a valid response to the item.

“Core0” shows the estimated mean using the initial sampling weights. “Cmbin\_Y” shows the unbiased mean.

“Bias0” biased and unbiased means when the initial sampling weights are used. Bias1 corresponds to item B4 (how districts report scores of students with disabilities).

Bias2 = B7a (technical assistance based on test scores of students with disabilities).

Bias3 = G1b (inservice training for general education teachers, row 68). Bias4 = B1b (content standards for math, rows 3–6).

Bias5 = G4 (placement of students in interim alternative settings, rows 75– 9). Bias6 = J2 (regular evaluation of parent satisfaction, row 65).

Bias7 = G1d (inservice training for other staff, row 70). Bias8 = B2e (performance standards in social studies, rows 39–42).

**Exhibit G2**

**Bias Estimates at Iteration 25**

Varno	Name	N	Core0	Cmbin_Y	Bias 0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
1	_TYPE_	0	0.0000	.	.	.	.	.	.	.	.	.	.
2	_FREQ_	405	40500.0000	.	.	.	.	.	.	.	.	.	.
3	CoMaNoS	271	3.4700	7.2600	-3.78720	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
4	CoMaSam	271	47.7100	45.3000	2.40830	0.00020	0.00020	0.00020	0.00000	0.00000	-0.00010	-0.00010	0.00010
5	CoMaSep	271	44.7400	37.1600	7.57520	0.00010	0.00010	0.00000	0.00000	0.00000	0.00010	0.00010	0.00010
6	CoMaNoD	271	4.0800	10.2800	-6.19630	-0.00030	-0.00030	-0.00030	0.00000	0.00000	0.00000	0.00000	-0.00020
7	CoRdNoS	271	3.1400	7.3500	-4.21360	-0.90430	-0.90430	-0.90430	-0.90430	-0.90430	-0.90430	-0.90430	-0.90430
8	CoRdSam	271	47.3600	44.2000	3.16460	0.87120	0.87120	0.87120	0.87100	0.87100	0.87100	0.87100	0.87120
9	CoRdSep	271	45.4200	38.1600	7.25530	0.04340	0.04340	0.04340	0.04330	0.04330	0.04340	0.04340	0.04340
10	CoRdNoD	271	4.0800	10.2800	-6.19630	-0.00030	-0.00030	-0.00030	0.00000	0.00000	0.00000	0.00000	-0.00020
11	CoScNoS	270	6.9000	8.1900	-1.29000	0.31990	0.31990	0.31990	0.31990	0.31990	0.31990	0.31990	0.31990
12	CoScSam	270	46.1100	42.3800	3.73460	1.99980	1.99980	1.99980	1.99970	1.99960	1.99960	1.99960	1.99980
13	CoScSep	270	42.8100	39.1200	3.69220	-2.46040	-2.46050	-2.46050	-2.46050	-2.46050	-2.46050	-2.46050	-2.46050
14	CoScNoD	270	4.1700	10.3100	-6.13690	0.14070	0.14070	0.14070	0.14100	0.14100	0.14100	0.14100	0.14080
15	CoSsNoS	270	7.3400	8.2500	-0.90880	1.09340	1.09340	1.09340	1.09330	1.09330	1.09330	1.09330	1.09340
16	CoSsSam	270	45.5400	42.3300	3.20720	0.31690	0.31690	0.31690	0.31680	0.31670	0.31670	0.31670	0.31690
17	CoSsSep	270	42.9800	39.1200	3.86070	-1.56700	-1.56710	-1.56710	-1.56710	-1.56710	-1.56710	-1.56710	-1.56710
18	CoSsNoD	270	4.1400	10.2900	-6.14910	0.16670	0.16670	0.16670	0.16700	0.16700	0.16700	0.16700	0.16680
19	CoWrNoS	270	5.7700	8.9200	-3.15070	-1.89950	-1.89950	-1.89950	-1.89950	-1.89950	-1.89950	-1.89950	-1.89950
20	CoWrSam	270	47.5900	41.6500	5.94240	3.29830	3.29830	3.29840	3.29820	3.29810	3.29810	3.29810	3.29830
21	CoWrSep	270	42.1400	39.1300	3.00760	-1.52680	-1.52680	-1.52680	-1.52690	-1.52690	-1.52680	-1.52680	-1.52680
22	CoWrNoD	270	4.5000	10.3100	-5.80930	0.11800	0.11800	0.11800	0.11820	0.11830	0.11820	0.11820	0.11800

**Exhibit G2**

**Bias Estimates at Iteration 25**

Varno	Name	N	Core0	Cmbin_Y	Bias 0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
23	CoFLNoS	268	23.6100	20.4600	3.15340	3.55690	3.55690	3.55690	3.55690	3.55680	3.55680	3.55680	3.55680
24	CoFLSam	268	42.0800	39.5400	2.54150	-2.56460	-2.56460	-2.56460	-2.56470	-2.56480	-2.56480	-2.56480	-2.56460
25	CoFLSep	268	27.1600	25.3300	1.83300	1.56870	1.56870	1.56870	1.56870	1.56870	1.56870	1.56870	1.56870
26	CoFLNoD	268	7.1400	14.6700	-7.52790	-2.56100	-2.56100	-2.56100	-2.56080	-2.56080	-2.56080	-2.56080	-2.56090
27	PrMaNoS	267	6.9000	10.0600	-3.16033	-1.12442	-1.12443	-1.12444	-1.12443	-1.12443	-1.12443	-1.12443	-1.12442
28	PrMaSam	267	39.5400	39.6700	-0.13438	2.02591	2.02592	2.02593	2.02587	2.02587	2.02587	2.02587	2.02592
29	PrMaSep	267	46.8000	44.4300	2.36606	-3.54675	-3.54673	-3.54674	-3.54692	-3.54692	-3.54692	-3.54692	-3.54675
30	PrMaNoD	267	6.7700	5.8400	0.92865	2.64525	2.64524	2.64524	2.64548	2.64548	2.64548	2.64548	2.64526
31	PrRdNoS	267	6.4000	10.0400	-3.63946	-1.97469	-1.97470	-1.97470	-1.97470	-1.97469	-1.97469	-1.97469	-1.97469
32	PrRdSam	267	39.5400	39.6100	-0.07438	2.08591	2.08592	2.08593	2.08587	2.08587	2.08587	2.08587	2.08592
33	PrRdSep	267	47.3000	44.5100	2.78520	-2.75648	-2.75647	-2.75647	-2.75665	-2.75666	-2.75666	-2.75666	-2.75649
34	PrRdNoD	267	6.7700	5.8400	0.92865	2.64525	2.64524	2.64524	2.64548	2.64548	2.64548	2.64548	2.64526
35	PrScNoS	266	9.9300	10.9100	-0.98464	-0.70800	-0.70801	-0.70801	-0.70801	-0.70801	-0.70801	-0.70801	-0.70801
36	PrScSam	266	37.8400	39.4000	-1.56221	1.44576	1.44578	1.44578	1.44573	1.44572	1.44573	1.44573	1.44578
37	PrScSep	266	45.0700	40.9800	4.09248	-0.75386	-0.75384	-0.75385	-0.75403	-0.75403	-0.75403	-0.75403	-0.75386
38	PrScNoD	266	7.1644	8.7100	-1.54560	0.01609	0.01608	0.01608	0.01631	0.01631	0.01631	0.01631	0.01609
39	PrSsNoS	267	11.9959	11.5539	0.44200	0.00001	0.00000	0.00000	-0.00001	0.00000	0.00000	0.00000	0.00000
40	PrSsSam	267	35.7027	38.8247	-3.12200	-0.00001	0.00000	0.00001	-0.00004	-0.00005	-0.00005	-0.00005	0.00000
41	PrSsSep	267	45.1961	40.9309	4.26520	0.00000	0.00001	0.00001	-0.00017	-0.00018	-0.00018	-0.00018	0.00000
42	PrSsNoD	267	7.1052	8.6905	-1.58530	0.00000	-0.00001	-0.00001	0.00023	0.00022	0.00022	0.00022	0.00000
43	PrWrNoS	266	7.4551	11.2300	-3.77490	-2.81609	-2.81611	-2.81611	-2.81611	-2.81610	-2.81610	-2.81610	-2.81610
44	PrWrSam	266	39.0345	38.6000	0.43450	2.75510	2.75511	2.75512	2.75506	2.75506	2.75506	2.75506	2.75511

**Exhibit G2**

**Bias Estimates at Iteration 25**

Varno	Name	N	Core0	Cmbin_Y	Bias 0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
45	PrWrSep	266	46.2396	41.5600	4.67960	0.06082	0.06084	0.06083	0.06065	0.06064	0.06064	0.06064	0.06081
46	PrWrNoD	266	7.2707	8.6100	-1.33930	0.00017	0.00016	0.00016	0.00039	0.00039	0.00039	0.00039	0.00017
47	PrFLNoS	267	28.2148	24.5200	3.69480	2.14942	2.14940	2.14940	2.14939	2.14940	2.14940	2.14940	2.14941
48	PrFLSam	267	31.7841	34.9900	-3.20590	-1.47005	-1.47005	-1.47005	-1.47008	-1.47009	-1.47008	-1.47008	-1.47003
49	PrFLSep	267	29.6240	28.5100	1.11400	0.74052	0.74055	0.74054	0.74040	0.74040	0.74040	0.74039	0.74052
50	PrFLNoD	267	10.3771	11.9900	-1.61290	-1.42989	-1.42990	-1.42989	-1.42971	-1.42971	-1.42971	-1.42971	-1.42990
51	Agged	288	20.2254	14.1300	6.09540	0.00000	0.00000	0.00000	0.00000	-0.00010	-0.00010	-0.00010	0.00010
52	Sep	288	5.9229	8.6500	-2.72710	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00010
53	Both	288	26.9533	29.0800	-2.12670	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00010
54	NotRep	288	1.6149	0.5600	1.05490	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
55	NoDWid	288	45.2835	47.5800	-2.29650	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00010	0.00010
56	sepboth	288	32.8762	37.7300	-4.85380	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	-0.00020
57	yb6a	286	44.6710	52.9300	-8.25900	-0.04270	-0.04280	-0.04290	-0.04290	-0.04290	-0.04290	-0.04290	-0.04270
58	yb6b	284	25.7611	33.4700	-7.70890	2.91060	2.91050	2.91050	2.91050	2.91050	2.91060	2.91060	2.91060
59	yb7a	279	32.2722	44.8800	-12.60780	0.00010	0.00000	0.00000	-0.00010	-0.00010	-0.00010	-0.00010	0.00020
60	yb7b	277	19.2304	28.5800	-9.34960	1.84720	1.84710	1.84710	1.84700	1.84700	1.84710	1.84710	1.84710
61	yh3a	273	16.2828	15.0500	1.23280	2.54650	2.54650	2.54640	2.54640	2.54640	2.54640	2.54640	2.54640
62	yh3b	271	7.1206	8.0500	-0.92940	0.97860	0.97860	0.97860	0.97860	0.97860	0.97860	0.97860	0.97860
63	yh4a	274	12.4742	12.3600	0.11420	2.36220	2.36220	2.36220	2.36220	2.36220	2.36220	2.36220	2.36220
64	yh4b	270	4.3290	6.8600	-2.53100	-0.46870	-0.46870	-0.46870	-0.46870	-0.46870	-0.46870	-0.46870	-0.46870
65	yj2	283	23.2959	46.7300	-23.43410	-0.00010	-0.00010	-0.00010	-0.00010	-0.00010	0.00000	0.00000	-0.00010
66	k3onepl	s 276	14.2981	17.5100	-3.21190	-2.84910	-2.84910	-2.84910	-2.84910	-2.84910	-2.84910	-2.84910	-2.84910

**Exhibit G2**

**Bias Estimates at Iteration 25**

Varno	Name	N	Core0	Cmbin_Y	Bias 0	Bias1	Bias2	Bias3	Bias4	Bias5	Bias6	Bias7	Bias8
67	nG1A	284	62.8625	70.9500	-8.08750	-1.05500	-1.05500	-1.05510	-1.05500	-1.05500	-1.05500	-1.05500	-1.05500
68	nG1B	282	64.7359	81.6400	-16.90410	0.00010	0.00000	0.00000	0.00000	0.00010	0.00010	0.00010	0.00000
69	nG1C	281	75.1852	87.2500	-12.06480	-0.66850	-0.66850	-0.66850	-0.66850	-0.66850	-0.66850	-0.66850	-0.66850
70	nG1D	282	59.2439	62.0000	-2.75610	0.00000	0.00000	-0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
71	nG1E	281	65.9008	73.0200	-7.11920	2.91530	2.91530	2.91520	2.91530	2.91530	2.91530	2.91530	2.91540
72	nG1F	278	33.7117	34.1600	-0.44830	5.62020	5.62020	5.62020	5.62020	5.62030	5.62030	5.62030	5.62030
73	nG1G	279	43.7070	45.7100	-2.00300	-0.14370	-0.14370	-0.14370	-0.14370	-0.14370	-0.14370	-0.14370	-0.14360
74	nG1H	281	50.6528	53.3900	-2.73720	-0.12170	-0.12170	-0.12170	-0.12170	-0.12170	-0.12170	-0.12170	-0.12180
75	Zero	242 4	3.6502	38.6900	4.96020	0.00020	0.00020	0.00020	0.00020	0.00000	0.00000	0.00000	0.00020
76	Gt0LEp5	242	6.7379	4.5900	2.14790	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
77	Gtp5LE2	242 1	9.3141	19.4400	-0.12590	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
78	Gt2LE5	242 1	5.8842	16.6900	-0.80580	0.00000	0.00000	0.00000	-0.00010	0.00000	0.00000	0.00000	0.00000
79	Gt5	242 1	4.4136	20.5900	-6.17640	-0.00020	-0.00020	-0.00020	-0.00010	0.00000	0.00000	0.00000	-0.00010

“NAME” shows the variable name. “N” is the number of Core districts that gave a valid response to the item.

“Core0” shows the estimated mean using the initial sampling weights. “Cmbin\_Y” shows the unbiased mean.

“Bias0” biased and unbiased means when the initial sampling weights are used. Bias1 corresponds to item B4 (how districts report scores of students with disabilities).

Bias2 = B7a (technical assistance based on test scores of students with disabilities).

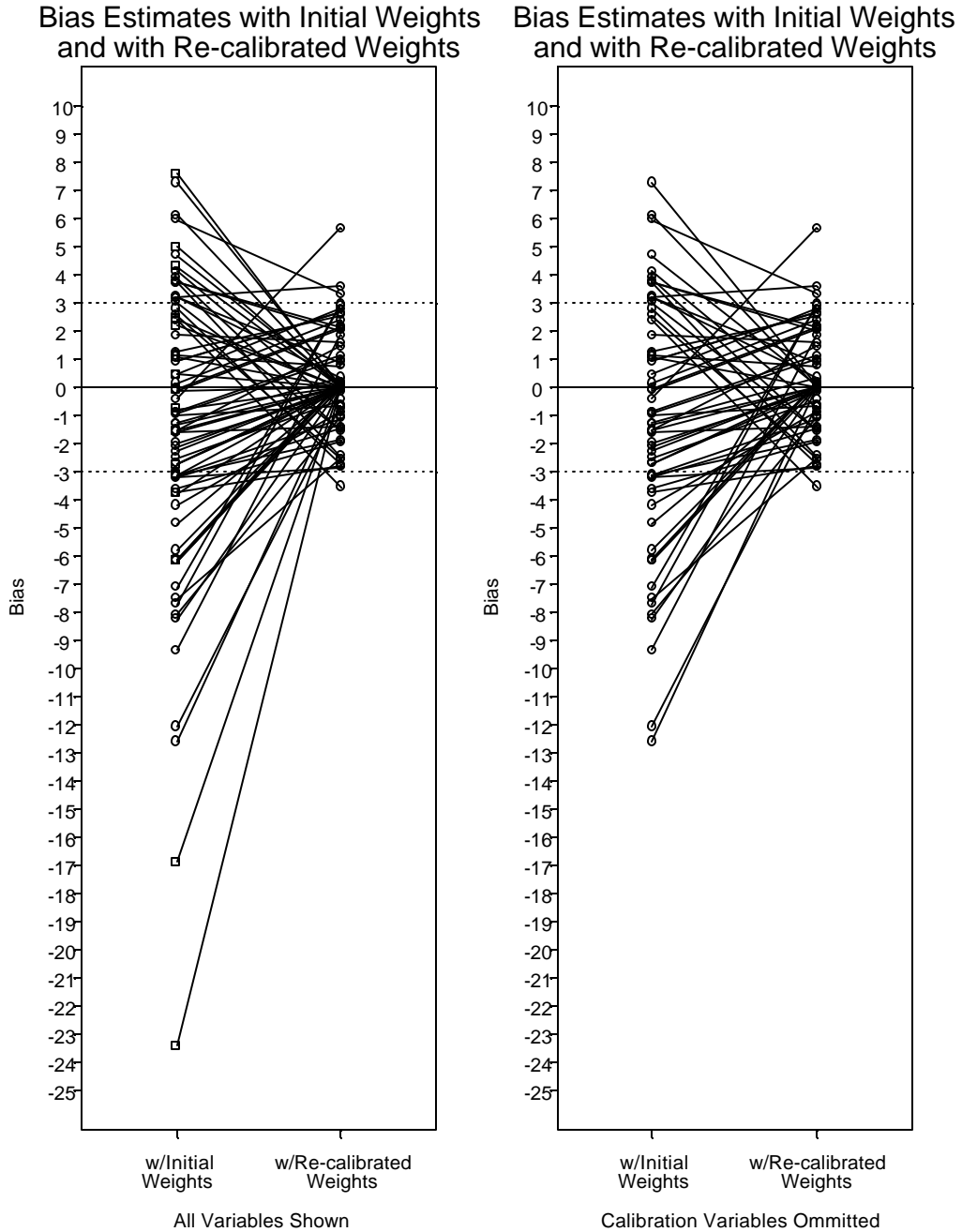
Bias3 = G1b (inservice training for general education teachers, row 68). Bias4 = B1b (content standards for math, rows 3–6).

Bias5 = G4 (placement of students in interim alternative settings, rows 75–79). Bias6 = J2 (regular evaluation of parent satisfaction, row 65).

Bias7 = G1d (inservice training for other staff, row 70). Bias8 = B2e (performance standards in social studies, rows 39–42).

**Exhibit G3**

**Bias Before and After Recalibrating Weights**



Calibration variables are shown with square plotting symbols. All other variables have round plotting symbols.

## A Note on Item Nonresponse

All of the reweighting discussed in this document was done in order to adjust for unit nonresponse. Unit nonresponse corresponds to a district that did not return a questionnaire at all (or completed so few items that their data were discarded). Item nonresponse describes the situation in which a district returned a questionnaire but did not provide an answer to a particular question. There are instances of item nonresponse for every item in the nonresponse survey. That is, there are no items that were answered by every one of the 298 Core districts. And there are no items that were answered by every one of the 107 NonPar districts. The calculation of the mean in the presence of item nonresponse is based on an implicit assumption that the reason for not responding to the item is not systematically related to the true answer to the question. This means, for example, that on a yes/no question, that if the true answers were known for districts that did not provide answers to the question, they would have the same probability of “Yes” responses as the districts that did provide answers. This is a strong assumption that, in truth, may or may not be reasonable. When there is very little item nonresponse, it is not terribly important whether or not the assumption is reasonable for the same reasons that bias due to associated unit nonresponse is unimportant when there is very little unit nonresponse. Because of the large amount of unit nonresponse, we have used methods to adjust the weights to minimize nonresponse bias. In cases where there is a large amount of item nonresponse for a particular survey item, we will attempt no similar types of adjustments. We will simply treat the data as unreliable and either not report the results or draw the reader’s attention to the problem.



# Appendix H: Items on the Nonresponse Surveys

## Items on the District Nonresponse Survey

Appendix F described the need for designing a nonresponse survey and the methodology for creating the survey. This appendix lists the questions that were included in the nonresponse survey.

The items that were in both the Core and the district nonparticipation surveys were the following: B1a–f, B2a–f, B3, B4, B6a–b, B7a–b, G1a–h, G4, H3a–b, H4a–b, J2, K3.

B1. What type of **content standards** does your district have for **students with disabilities** in each of the following subjects? Content standards describe what every student should know and be able to do in the core academic content areas. Please check **one** box on each line.

	No district standards for any students	Same district standards for all students	Separate district standards for students with disabilities	Modified or enhanced district standards for students with disabilities	No decision made about district standards for students with disabilities
	(1)	(2)	(3)	(4)	(5)
a. Foreign language ....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Mathematics .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Reading .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Science .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Social studies .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Writing .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

B2. What type of **performance standards** does your district have for **students with disabilities** in each of the following subjects? Performance standards answer the question, “How good is good enough?” They gauge the degree to which content standards have been attained by individuals or groups, indicating the nature of the evidence and quality of student performance on agreed upon tasks or measurement instruments. Please check **one** box on each line.

	No district standards for any students	Same district standards for all students	Separate district standards for students with disabilities	Modified or enhanced district standards for students with disabilities	No decision made about district standards for students with disabilities
	(1)	(2)	(3)	(4)	(5)
a. Foreign language ....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Mathematics .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Reading .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

	No district standards for any students	Same district standards for all students	Separate district standards for students with disabilities	Modified or enhanced district standards for students with disabilities	No decision made about district standards for students with disabilities
d. Science.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Social studies .....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Writing.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

B3. Do you have any **district-wide** assessments not required by the state?

- 1)  Yes
- 2)  No (*Go to B5*)

B4. How does your district report the scores of **students with disabilities** who participate in **district-wide** assessments? Please check **one** box.

- 1)  Aggregated with the scores of students without disabilities
- 2)  Reported separately from those of students without disabilities
- 3)  Reported both aggregated and separately
- 4)  Not reported

B6. Does your district provide additional technical or financial assistance to schools that do poorly on tests measuring student achievement? Please check **one** box on each line.

	Yes, we provide additional assistance (1)	No, we don't provide additional assistance (2)
a. Technical assistance .....	<input type="checkbox"/>	<input type="checkbox"/>
b. Financial assistance .....	<input type="checkbox"/>	<input type="checkbox"/>

B7. Does your district provide additional technical or financial assistance **focusing on students with disabilities** to schools where **students with disabilities** do poorly on tests measuring student achievement? Please check **one** box on each line.

	Yes, we provide additional assistance (1)	No, we don't provide additional assistance (2)
a. Technical assistance .....	<input type="checkbox"/>	<input type="checkbox"/>
b. Financial assistance .....	<input type="checkbox"/>	<input type="checkbox"/>

G1. Does your district provide any of the following types of support to schools that are developing school-wide approaches to discipline? Please check one box on each line.

	<b>Yes</b> (1)	<b>No</b> (2)
a. Inservice training for administrators.....	<input type="checkbox"/>	<input type="checkbox"/>
b. Inservice training for general education teachers.	<input type="checkbox"/>	<input type="checkbox"/>
c. Inservice training for special education teachers .	<input type="checkbox"/>	<input type="checkbox"/>
d. Inservice training for other school staff.....	<input type="checkbox"/>	<input type="checkbox"/>
e. Guidance or consultation from district staff .....	<input type="checkbox"/>	<input type="checkbox"/>
f. Funds for additional security or other staff.....	<input type="checkbox"/>	<input type="checkbox"/>
g. Program for developing partnerships with local community and service agencies.....	<input type="checkbox"/>	<input type="checkbox"/>
h. Help developing partnerships with parents/guardians.....	<input type="checkbox"/>	<input type="checkbox"/>

G4. During the past year, how many **students with disabilities** has your district placed in interim alternative placements at least once because of behavior problems?

Number of students \_\_\_\_\_

H3. Does your district provide additional technical or financial assistance to schools with high dropout or low graduation rates? Please check **one** box on each line.

	<b>Yes, we provide additional assistance</b> (1)	<b>No, we don't provide additional assistance</b> (2)
a. Technical assistance .....	<input type="checkbox"/>	<input type="checkbox"/>
b. Financial assistance .....	<input type="checkbox"/>	<input type="checkbox"/>

H4. Does your district provide additional technical or financial assistance focusing on **students with disabilities** to schools where **students with disabilities** have high dropout or low graduation rates? Please check **one** box on each line.

	<b>Yes, we provide additional assistance</b> (1)	<b>No, we don't provide additional assistance</b> (2)
a. Technical assistance .....	<input type="checkbox"/>	<input type="checkbox"/>
b. Financial assistance .....	<input type="checkbox"/>	<input type="checkbox"/>

J2. Does your district regularly evaluate parent/guardian satisfaction with **special education** services?

1)  Yes

2)  No (*Go to J6*)

K3. During the past year, how many cases involved mediation, impartial due process, state administrative review of the hearing decision or state/federal judicial review of hearing decision? Count each case only **once**, even if it involved more than one type of dispute.

Number of cases \_\_\_\_\_

## Items on the School Nonresponse Survey

# SLIIDEA 1999–2000 NONRESPONSE SURVEY SCHOOL QUESTIONNAIRE

According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB number. The valid OMB control number of this information collection is 1820-0633. The time required to complete this information collection is estimated to average 10 minutes, including the time to review instructions, search existing data resources, gather the data needed and complete and review the information collection. If you have any comments concerning the accuracy of the time estimate(s) or suggestions for improving this form, please write to: U.S. Department of Education, Washington, DC 20202-4651. If you have comments or concerns regarding the status of your individual submission of this form, write directly to: Office of Special Education Programs (OSEP), 400 Independence Avenue SW, Washington, D.C. 20202-8971.

**A project of the Office of Special Education Programs in the U.S. Department of Education.** This survey is authorized by law [P.L. 105-17, Section 674(b)]. While an individual's participation in this study is voluntary, your cooperation is needed to make the results of this study comprehensive, accurate, and timely.

OMB No. 1820-0633 Approval Expires: 04/30/2003

School ID: «SCHID»

School Name: «School\_Name»

---

---

Throughout this questionnaire, **students with disabilities** means students with IEPs who are receiving special education and/or related services.

---

---

### SCHOOL CHARACTERISTICS

A6. During the 1999–2000 school year, what was the total number of students enrolled in this school?

Total number of students \_\_\_\_\_

A10. During the 1999–2000 school year, how many **special education students** were enrolled in this school? Only include students with IEPs in your answer.

Total number of special education students \_\_\_\_\_

---

---

**ASSESSMENT**

B7. During the 1999–2000 school year, did your school provide information to parents/guardians about the aggregate performance of students in your school on state, district or standardized tests measuring student achievement?

- 1)  Yes  
2)  No

B8. During the 1999–2000 school year, did your school provide information to parents/guardians of **special education students** about the aggregate performance of **special education students** in your school on state, district or standardized tests measuring student achievement?

- 1)  Yes  
2)  No
- 
- 

**BEHAVIOR ISSUES AND DISCIPLINE**

F1. During the 1999–2000 school year, did you have a school-wide behavior plan or discipline policy that included the following? Please check **one** box on each line.

	<b>Yes</b> (1)	<b>No</b> (2)
a. A list of infractions for which students are automatically recommended for expulsion .....	<input type="checkbox"/>	<input type="checkbox"/>
b. A list of infractions for which students are automatically suspended .....	<input type="checkbox"/>	<input type="checkbox"/>
c. Description of positively stated behavior expectations .....	<input type="checkbox"/>	<input type="checkbox"/>
d. Instruction for students in how to solve problems with peers .....	<input type="checkbox"/>	<input type="checkbox"/>
e. Uniform school-wide consequences for disruptive behavior .....	<input type="checkbox"/>	<input type="checkbox"/>
f. Uniform school-wide rewards or incentives for positive or good behavior .....	<input type="checkbox"/>	<input type="checkbox"/>

F4. During the 1999–2000 school year, did your school have an **in-school** suspension program?

- 1)  Yes  
2)  No (*Go to F9*)

F5. Did your **in-school** suspension program include any of the following? Please check **one** box on each line.

	<b>Yes</b> (1)	<b>No</b> (2)
a. Counseling and/or support designed to prevent behavior from recurring .....	<input type="checkbox"/>	<input type="checkbox"/>
b. Consistent staffing by the same staff member ...	<input type="checkbox"/>	<input type="checkbox"/>
c. Staffing by rotating staff members .....	<input type="checkbox"/>	<input type="checkbox"/>
d. Academic support to help students with assignments from their regular classes .....	<input type="checkbox"/>	<input type="checkbox"/>

F6. During the 1999–2000 school year, how many students **with** and **without** disabilities received **in-school** suspension? If a student received **in-school** suspension more than once, please count that student one time only.

Number of students **with** disabilities \_\_\_\_\_

Number of students **without** disabilities \_\_\_\_\_

F9. Thinking about the 1999–2000 school year, generally, do you believe your school was able to adequately and appropriately discipline students with disabilities who presented challenging, disruptive and/or dangerous behaviors?

- 1)  Yes  
 2)  No

### PARENTS/GUARDIANS

H2. During the 1999–2000 school year, did your school have parent workshops, training, support groups or discussion groups to address issues relating to students with disabilities? Please check **one** box on each line.

	<b>Yes</b> (1)	<b>No</b> (2)
a. Parent workshop or training on issues related to students with disabilities .....	<input type="checkbox"/>	<input type="checkbox"/>
b. Support and/or discussion group for parents/guardians of students with disabilities ...	<input type="checkbox"/>	<input type="checkbox"/>

H3. During the 1999–2000 school year, did you receive reports from your state or district about the level of parent/guardian satisfaction with **special education** services in your school?

- 1)  Yes  
2)  No

---

---

### DISPUTE RESOLUTION

I3. During the 1999–2000 school year, how many cases involved mediation, impartial due process, state administrative review of the hearing decision (if applicable in your state) or state/federal judicial review of hearing decision? Count each case only **once**, even if it involved more than one type of dispute.

Number of cases total \_\_\_\_\_

THANK YOU!

# Appendix I: Statistical Methods

## Overview

This appendix provides an account of the statistical methods used in the study. Almost all outcomes were binary and were modeled by logistic regressions using the LOGISTIC procedure in SUDAAN®. This correctly accounts for the fact that the data were obtained via a complex sample rather than a simple random sample.

The purpose of the modeling exercise was to explain (or account for) aspects of district-level and school-level behavior. District behavior was thought to depend on predictors at the district and state levels, while school behavior was thought to depend on predictors at the school, district and state levels. Predictors in the final model were then selected from the preliminary list of plausible candidate predictors. The purpose of this kind of selection was to discard unimportant variables and retain a parsimonious model purporting to describe the particular district or school behavior.

For example, our preliminary model accounting for whether a district provided assistance to schools based on high dropout rates included all district demographic variables (size, percent minorities, percent free lunch, percent IEP, degree of urbanization), the state demographic variable (region), and several state policies (e.g., whether the state rewarded or sanctioned districts based on high dropout rates, and whether the state itself provided assistance based on high dropout rates). The predictors selected for the final model were district size and the percent of minorities in the district.

## Data

### Groups of Variables

Variables used in the logistic models fell into six groups, being either demographic variables or outcomes at the school, district or state levels. These are outlined in Exhibit I1.

<b>Exhibit I1</b>			
<b>Groups of Variables Used in Models</b>			
	<b>School</b>	<b>District</b>	<b>State</b>
<b>Demographic Variables</b>	Size Percent free lunch Percent minorities Percent IEP Urbanicity	Size Percent free lunch Percent minorities Percent IEP Urbanicity	Region
<b>Outcomes</b>	e.g., School has a formal dropout-prevention program	e.g., District provides assistance to schools based on high dropout rates	e.g., State provides assistance to schools based on high dropout rates

There were five school and district demographic variables, and a single state demographic variable representing region (Midwest, Northeast, South and West). School and district demographic variables were obtained either from their respective surveys or from the NCES Common Core of Data. District variables were district size, the percent of minorities in the district, the percent of students in the district receiving free lunch, the percent of students in the district with Individual Educational Plans (IEP) and the urban status of the district (rural, suburban, urban). School demographic variables were the school analogues of these district variables: school size, the percent of minorities in the school, the percent of students receiving free lunch in the school, the percent of students in the school with IEPs and the school’s urban status.

For models explaining district behavior, the response variable was a district outcome, and predictors were district and state demographic variables, as well as relevant district and state policy variables. In the case of models explaining school behavior, the response variable was a school outcome, and predictors came from all six groups: school, district and state demographic variables, and relevant school, district and state policy variables. However, for the school models we used only two district demographic variables—district size and the percent of students in the district with IEPs—because of the high correlations between other district variables and their school analogues. The main diagonal of Exhibit I2 shows these correlations—or more precisely, the squared correlations ( $R^2$  statistics). In particular the school-district squared correlations for percent minorities, urbanicity and percent free lunch were 75 percent, 63 percent and 45 percent, respectively. The point here is that the data for percent minorities in schools are very similar to the data for percent minorities in districts, and, as a result, it is difficult to distinguish their individual effects on a particular school outcome.

<b>Exhibit I2</b>					
<b>Signed Squared Correlation Between School and District Demographic Variables</b>					
	<b>School Size</b>	<b>School Free Lunch %</b>	<b>School Minority %</b>	<b>School IEP %</b>	<b>School Urbanicity</b>
<b>District Size</b>	11	2	16	-3	39
<b>District Free Lunch %</b>	0	45	34	0	4
<b>District Minority %</b>	5	21	75	-1	31
<b>District IEP %</b>	0	0	0	13	0
<b>District Urbanicity</b>	7	1	18	-2	63

### **School, District and State Outcomes**

Whether a school or district outcome is considered a “response variable” or “policy variable” depends on the modeling context. Conceptually at least, the idea is that a policy variable influences a response variable rather than the other way around. In any event, school, district and state outcomes were obtained from the school, district and state surveys. These were frequently answers to binary survey questions such as “Did the district provide financial assistance to schools based on high dropout rates?” and “Did the district provide technical assistance to schools based on high dropout rates?” In other cases, the binary outcome was a composition of answers to several questions. For example, although the question “Did the district provide either financial or technical assistance to schools based on high dropout rates?” was not asked on the survey, it could readily be deduced from the above two questions that were asked.

## Continuous, Ordinal and Categorical Variables

The variable representing region is categorical, while all school and district demographic variables are either ordinal (e.g., urbanicity) or continuous (e.g., size). For reasons of robustness and uniformity, all ordinal and continuous variables were grouped into categories with three or four levels and assigned integer scores. For example, the number of students in a school were categorized into three levels ( $\leq 250$ , 250–750,  $> 750$ ) and then assigned integer scores (1, 2, 3). The correspondence between integer and original scores is displayed in Exhibit I3.

Exhibit I3		
Demographic Variables: Integer and Ordinal Scores		
	School	District
Size	(1,2,3) = ( $\leq 250$ , 250–750, $> 750$ )	(1,2,3) = ( $\leq 2000$ , 2000–8000, $> 8000$ )
Percent Free Lunch	(1,2,3) = ( $\leq 25\%$ , 25–50%, $> 50\%$ )	(1,2,3) = ( $\leq 25\%$ , 25–50%, $> 50\%$ )
Percent Minorities	(1,2,3) = ( $\leq 25\%$ , 25–75%, $> 75\%$ )	(1,2,3) = ( $\leq 25\%$ , 25–75%, $> 75\%$ )
Percent IEP	(1,2,3,4) = ( $\leq 8\%$ , 8–15%, 15–25%, $> 25\%$ )	(1,2,3) = ( $\leq 8\%$ , 8–15%, $> 15\%$ )
Urbanicity	(1,2,3) = (rural, suburban, urban)	(1,2,3) = (rural, suburban, urban)

Although all demographic variables were given numerical values, not all were actually treated as “continuous” predictors in a particular model. Rather, the choice between continuous and categorical treatment hinged on the outcome of a statistical test. If the observed data were sufficiently far from linear ( $p < 0.025$  for a test on the irrelevance of the quadratic parameter), the predictor was treated categorically, and the values (1, 2, 3) were merely convenient labels for the three categories. Otherwise the predictor was treated continuously and in a linear way, in the sense that a unit increase in the predictor was associated with a constant change in the predicted response, in this case the log (odds). This exercise was carried out for every model, and the outcome of the continuous-categorical test was, of course, specific to a given model. Thus district size might be treated continuously for one district outcome and categorically for another.

## Imputed Demographic Data

In cases where demographic variables were missing, imputed values were used. In the absence of such imputation, schools and districts with any missing data would be discarded. For districts, one value for percent IEP was missing, along with two values for percent minorities and 27 values for percent free lunch. Of the 298 sampled districts, 29 were missing at least one demographic value (one district was missing both percent minorities and percent free lunch); so 269 districts had complete data. For schools, two values for percent minorities and 13 values for percent free lunch were missing. Of the 744 sampled schools, 15 were missing a demographic value; so 729 schools had complete data.

Imputation was based on a random draw from the predictive distribution of a missing demographic variable. This entailed fitting a multinomial regression with the missing demographic variable as the multinomial response, and the remaining demographic variables, and region, as the predictors. Rather than impute the missing demographic variable by its mean, we took a random draw from a multinomial distribution with this mean, thus maintaining the appropriate amount of uncertainty in

the imputation. Because several demographic variables were missing, the imputation scheme was carried out in order of missingness. For example, in the district case, we imputed percent IEP (one missing), then percent minorities (two missing) and finally percent free lunch (27 missing). This served as an approximation to a fully iterative procedure that would continue until convergence was obtained.

For example, the first step of the imputation scheme for districts went as follows. The multinomial model had percent IEP as a three-category-response variable. The predictors were the remaining district demographic variables (size, percent minorities, percent free lunch and urbanicity), and region. This model was fit to the 269 districts with complete data, and predicted probabilities were obtained for the percent IEP categories of the missing district. The missing district in this case was a large urban district in the Northeast with mid-level percent minorities and low-level percent free lunch. With these covariate values, the predicted probabilities for the three-category-response variable were  $(p_1, p_2, p_3) = (0.004, 0.629, 0.367)$ . A random draw from a multinomial with these probabilities produced the second category (the most likely category in this case), and this was used as the imputed value for percent IEP for the missing district.

### Sampling Zeros

Sparse data cause serious problems for logistic modeling. We describe our solution to this problem by way of an example. The third column of Exhibit I4 shows the number of districts in all response-predictor combinations for a particular response variable and a particular predictor. The predictor was the percent of minorities in the district, and the response variable indicated whether the district provided regular education for the majority of students with mental retardation (1 = Yes, 0 = No). Thus 12/141 districts with less than 25 percent minorities provided regular education for the majority of students with mental retardation, while the numbers for the middle- and high-percent-minority categories were 6/72 and 0/19, respectively. The three observed proportions of districts providing the service were therefore (0.09, 0.08, 0), and the observed odds were (0.09, 0.09, 0).

<b>Exhibit I4</b>			
<b>Original and Modified Counts for a District Predictor and District Outcome</b>			
<b>Response</b>	<b>Predictor</b>	<b>Original Counts</b>	<b>Modified Counts</b>
0	1	129	129
1	1	12	12
0	2	66	66
1	2	6	6
0	3	19	18
1	3	0	1

The fact that all districts in the sample had the same outcome for certain predictor levels made it impossible to fit sensible logistic models. In the current example this arises because, in terms of ratios of odds, districts with high percent minorities are infinitely less likely to provide the service compared to other districts. In such circumstances, the predictor is often either dropped or collapsed into fewer levels. Instead, we viewed the problem as one involving a sampling zero. In the above example, although the proportions of districts providing the service were *actually* (12/141, 6/72, 0/19), a slightly different realization could have given (12/141, 6/72, 1/19), and this is the one we used. This minor modification (using the counts in column four rather than those in column three) solved the technical problem in a way that uses the available sample information. In this case the data

suggest that the three percent minority levels were similar with respect to this outcome, and this is borne out by the modified proportions, (0.09, 0.08, 0.05). The above discussion has implicitly assumed data with equal weights. To accommodate the unequal weights in our sample, we used the average weight for all districts in the offending level (the highest percent minority level), which in this case was 46.2.

## Logistic Models

Almost all models we considered were weighted logistic models based on the sampling weights described in earlier appendices. These models were fit using the LOGISTIC procedure in SUDAAN®, which correctly accounts for the fact that the data were obtained via a complex sample rather than a simple random sample. Standard software is inappropriate because it assumes a simple random sample. This amounts to assuming that (for large samples) the sample distribution looks like the population distribution, which is unlikely to be the case under our sampling scheme.

The response variable was a binary district or school outcome ( $y_i = 0$  or  $y_i = 1$  for the  $i^{\text{th}}$  district), and the predictors were treated continuously or categorically as outlined previously. The logistic model purports to describe a particular district behavior, or rather the probability of this behavior  $p_i = \text{prob}(y_i = 1)$ , as a function of the predictors. For example, if a particular district behavior was thought to depend on district demographics, region and several state policies, then the probability that a district exhibits this behavior is

$$p_i = \exp(\alpha + \beta'd_i + \gamma'r_i + \delta's_i) / (1 + \exp(\alpha + \beta'd_i + \gamma'r_i + \delta's_i))$$

where  $p_i$  is probability that the  $i^{\text{th}}$  district provides assistance,  $d_i$ ,  $r_i$ , and  $s_i$  are predictor vectors for district demographics, region, and state policies, and  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  are their corresponding vector parameters.

The form of the logistic function restricts the probability to the interval between zero and 1, so that as a predictor (with a positive effect) increases,  $p_i$  also increases, but in a nonlinear S-shaped fashion. This is how it should be, but it also means that the effect of a predictor depends on the specific level of the predictor. For example, the effect of increasing a predictor from 1 to 2 has a different effect on  $p_i$  than an increase from 2 to 3. Although it is possible to avoid this awkwardness by working with odds rather than probabilities (the effect of a predictor has a constant multiplicative effect on the odds), we typically expressed results in terms of probabilities and proportions. If the model contained a single predictor—for example, size—we expressed results in terms of the proportion of districts exhibiting the behavior at each level of the predictor (in small, medium and large districts). If the model contained several predictors—for example, size and percent minorities—then the proportions for one variable (e.g., size) were obtained at the population average of the other predictors (percent minorities).

### Model Selection

The way in which we selected predictors paid particular attention to the correlation among predictors. As previously noted, variables fell into six groups for models describing school behavior and four groups for models describing district behavior. In this study, the correlation was typically low *between* the variable groups and high *within* these groups. For district models, for example, the

correlation between percent minorities and percent free lunch was over 0.7, the correlation between two state predictors sometimes reached 0.9, but the correlation between a district demographic variable and a state policy variable was typically less than 0.2.

Our selection process exploited this correlation pattern by first selecting among the three groups, and then selecting among individual variables within the selected groups. For example, if the district demographic group was the only group selected at the first stage, then the second stage focused exclusively on selecting individual district demographic variables and ignored region and state policies. The virtue of the two-stage method is that selection at the first stage was unaffected by the high correlations within the group, and this afforded high-powered tests for group effects. Of course, high within-group correlations remained and needed to be confronted at the second stage, but the first stage at least provided us with the knowledge that *some* predictor in a selected group was important.

For the purpose of exposition, suppose that percent minorities and percent free lunch were the only variables in the district demographic group, and that this group was selected at the first stage ( $p = 0.001$ ). Because of the high positive correlation between the two individual variables, we can expect nonsignificant individual  $p$ -values. Suppose these were  $p = 0.6$  for percent minorities and  $p = 0.7$  for percent free lunch. We know that either one or both of the variables are important because the joint test was highly significant, but without a very large sample size it is not possible to determine which, because of the loss of power induced by the high correlation between the two variables. The loss of power in this situation is unavoidable and calls for an alternative selection criterion.

Our alternative was based on choosing the model with the lower mean squared error of the parameter estimates. This combines the decreased variance with the increased bias that results from dropping one of the predictors. The aim is to contrast the situation in which *both* correlated variables are important with one in which *one* of the correlated variables (percent minorities, say) is important. In both cases, dropping the least significant variable (percent free lunch with  $p = 0.7$ ) will decrease the variance of the parameter estimates. However, in the first case the increased bias will tend to outweigh the reduced variance, while in the second case it will not. Thus in the first case, percent free lunch will be kept, and in the second case it will be dropped.

