WORKING
PAPER

Supplemental Analyses of ISBA Survey Responses

SCOTT NAFTEL, LAURA S. HAMILTON, AND
BRIAN M. STECHER

RAND EDUCATION

PREFACE

Researchers from the RAND Corporation have been conducting the Implementing Standards-Based Accountability (ISBA) project, funded by the National Science Foundation, to examine the implementation of the NCLB standards-based accountability provisions in three states. Findings from this study have been reported in two documents, *Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States,* MG-589-NSF; and *Pain and Gain: Implementing No Child Left Behind in Three States, 2004 to 2006*, MG-784-NSF.

Staff members from the Government Accountability Office (GAO) have expressed interest in using analyses of teacher survey data from the ISBA study in a report they are preparing to examine the effects of state accountability testing on instruction. Many of the specific analyses in which GAO was interested were not included in the published RAND reports. GAO staff sent RAND a list of requested analyses, and RAND staff produced this report in response to their request.

This report reproduces key sections of the earlier reports dealing with the development of the surveys, the sampling of teachers, and the application of non-response and sampling weights. The published reports also discuss the limitations of the study, some of which stem from the self-report nature of the survey data. Users of these analyses should keep these limitations in mind and should apply appropriate caution when interpreting these results.

This research was conducted by RAND Education, a unit of the RAND Corporation. It is part of a larger body of RAND Education work addressing assessment and accountability. It was sponsored by the National Science Foundation under grant number REC-0228295. Any opinions, findings, conclusions, or recommendations expressed in this monograph are those of the authors and do not necessarily reflect the views of the National Science Foundation.

CONTENTS

TABLES

1. INTRODUCTION

This working paper was produced in response to a request from the Government Accountability Office for additional analyses of survey responses from the Implementing Standards Based Accountability study. The report consists primarily of tables of results from computer analyses. To provide some useful context, we begin with excerpts from earlier reports describing the focus of the study and the sampling, data collection and analysis methods.

2. PURPOSE

The No Child Left Behind Act (NCLB) of 2001 (20 U.S.C. § 6311 et seq.) is arguably the primary policy initiative affecting schools and districts in the United States today, and its standards-based accountability (SBA) provisions are perhaps its most potent component. NCLB requires states[1] to adopt content and achievement standards, to measure student progress toward those standards, and to implement a series of interventions and sanctions in schools and districts that fail to meet their targets. Together, these standards, assessments, and consequences constitute an SBA system. Since 2001–2002, each of the states has been developing and implementing such a system that is consistent with NCLB yet unique to the state. Recent research suggests that these systems have already had a large impact on schools and districts.

The Implementing Standards-Based Accountability (ISBA) study was designed to identify factors that enhance the implementation of SBA systems, foster changes in school and classroom practice, and promote improved student achievement. Advocates of SBA often claim that these policies will encourage educators to focus their efforts on improving achievement in core subjects and boost student proficiency, whereas detractors worry that attaching higher stakes to test results will lead to adverse consequences, such as narrowing of the curriculum or excessive test preparation resulting in invalid test scores. Whether NCLB's accountability provisions ultimately turn out to be beneficial or detrimental is likely to depend in large part on how states, districts, and schools implement the provisions and respond to the incentives.

The ISBA study gathered data from selected states, districts, schools, and teachers longitudinally for three years to address three broad questions:

---

[1] The term "state" is used in this paper to refer to all the jurisdictions that are responsible for implementing NCLB, which include the 50 states, the District of Columbia, and Puerto Rico.

- What strategies are used in implementing SBA at the state, district, and school levels?
- Which state, district, and school implementation strategies are associated with changes in classroom practice?
- Which features of SBA implementation are associated with student attainment of academic standards?

Answers to these questions are intended to help policymakers, educators, and researchers understand the ways in which SBA policies are implemented at all levels of the system; explain relationships among implementation, instructional practices, and student achievement; and identify successful practices that can be promulgated to make SBA function more effectively. This working paper supplements previously published monographs to provide additional evidence related to the first research question.

3. SAMPLING

The three states we chose to study were selected to provide diversity in terms of geography, demography, and their approaches to implementing NCLB (see Appendix A, Tables A.1 and A.2). California is the largest of the states with over 6.4 million students in 2005-2006 and has the most diverse student population. It is the only state of the three with a large population of English language learners (about one-quarter of the student population in 2005-2006) and has much larger populations of Hispanic and Asian students than the other states (47 percent and 11 percent of the student population, respectively). Georgia has the largest proportion of African American students (38 percent in 2005-2006). Pennsylvania is the least diverse of the states and has the lowest percentage of economically disadvantaged students; less than a third of Pennsylvania students were eligible for free or reduced-price lunches in 2005-2006, as compared to about half of students in California and Georgia. Of the three states, California had the most complete SBA system prior to the enactment of NCLB. Georgia had just started to implement an SBA system when NCLB was enacted and therefore had the smoothest transition to NCLB compliance. Pennsylvania had a strong tradition of local control and, therefore, had to work more quickly than the other states to develop the standards, assessments, and accountability polices that NCLB requires.

In the first year of our data collection (2003-2004), we selected an initial sample of 27 districts in each state. Districts in each state were stratified based on the number of elementary and middle schools, and a random sample was drawn within each stratum. (The technical details of the sampling design and weighting are provided in Appendix B.) We were not able to recruit as many districts as we had hoped for and, therefore, drew a supplemental sample of 23 districts. The total sample for the first year was 104 districts, 68 of which agreed to participate in the study, representing a cooperation rate of 65 percent. In 2004-2005, we selected an additional supplemental sample of 28 districts in order to yield greater analytic power and to increase

the number of districts with high percentages of schools struggling to meet NCLB requirements. This increased the total sample to 132 districts, 92 of which agreed to participate in the 2004-2005 year of the study. All 92 of these districts agreed to continue with the study for the 2005-2006 year of data collection (see Appendix A, Table A.3).

The school sample was restricted to include only "regular" public schools; charter schools, alternative schools, vocational schools, special education schools, and small schools were all excluded. In 2003-2004, 297 schools were randomly selected from the cooperating districts. The number of schools sampled in each district was based on district size and ranged from one to five elementary schools, and one to five middle schools. Of these schools, 267 agreed to participate in 2003-2004, representing a cooperation rate of 90 percent. The participating schools were contacted again in 2004-2005 as well as additional schools from the supplemental sample of districts. The total sample for 2004-2005 was 353 schools, 301 of which participated, representing a cooperation rate of 85 percent. In 2005-2006, two schools in Pennsylvania dropped out of the study, decreasing the cooperation rate by approximately 0.5 percent (see Appendix A, Table A.4). Most of the schools in the sample made AYP each year of the study, but the rates were different across the states. For example, in 2006, 61 percent of the sampled schools in California, 76 percent of the sampled schools in Georgia, and 83 percent of the sampled schools in Pennsylvania made AYP, according to principal reports. Principals also reported that only a small fraction of the schools in the sample were identified for improvement 15, 23, and 10 percent, respectively, in California, Georgia, and Pennsylvania.

Each year, we asked cooperating elementary schools for a roster of teachers who taught math and science in grades three, four, and five, and we asked middle schools for a roster of teachers who taught these subjects in grades seven and eight. We administered surveys to all of those teachers. Each annual teacher sample was drawn independently; we did not track teachers over time. Over 3,000 teachers were surveyed each year.

4. DATA COLLECTION

At the state level, we conducted semi-structured face-to-face interviews with a variety of key stakeholders, including high-level state department of education officials, legislators and legislative staff, state board of education staff, and union and state school boards' association leaders. We also collected relevant documents, such as copies of state content standards. Most of the state-level data collection occurred in the fall of 2003.

District level data were collected from superintendents through paper-and-pencil surveys in January and February of 2004, 2005, and 2006, and through semi-structured telephone interviews in the spring of 2004 and 2006. School level data were gathered each year through principal and teacher surveys, and annual case studies were conducted in a small subsample of schools. Many of the survey questions appeared in both the principal and teacher surveys, allowing us to compare responses between these groups. Principal and teacher surveys were distributed in January and February of each year, and responses were collected through June. Each survey instrument (superintendent, principal, and teacher) was pilot tested with representatives from the appropriate respondent group using structured "think aloud" cognitive interviews.

Response rates were quite high for each of the three surveys each year. The survey response rate for the superintendent survey was 88 percent in 2003–2004, and 73 percent in 2004–2005 and 2005–2006. The response rate for the principal survey was between 85 and 87 percent each year. The response rate for the teacher survey was 83 percent in 2003–2004, 87 percent in 2004–2005, and 82 percent in 2005–2006 (see Appendix, A, Tables A.5, and A.6).

We also conducted annual case study visits to two elementary schools and one middle school in two districts in each state. We interviewed principals, teachers, and other staff, and conducted parent focus groups when principals agreed.

5. SURVEY ANALYSES

To analyze survey responses, we generated state-specific sampling and nonresponse weights for each state. Using these weights, we are able to report statewide estimates of the responses of superintendents, principals, and teachers from regular public schools and districts. Because we excluded some schools that are subject to NCLB requirements but that operate outside a traditional district governance structure, such as charter schools, all of the results generalize only to regular public schools in the respective states. One of the consequences of our sampling strategy in which teachers and principals are nested within schools and schools are nested within districts is that the number of responses grows progressively smaller as we move from teachers to principals. As a result, the summary statistics based on teacher responses are more precise than those based on principal responses. To help the reader interpret the results, we include estimates of the standard errors (SEs) associated with the survey responses in all of the tables in the working paper.

6. RESULTS

As noted earlier, the primary results from our analyses have been presented in two RAND monographs. The results of the new analyses of the ISBA survey data can be downloaded from http://www.rand.org/pubs/working_papers/WR628/.

REFERENCES

Hamilton, L. S., B. M. Stecher, J. Marsh, J. S. McCombs, A. Robyn, J. L. Russell, S. Naftel, and H. Barney (2007). *Implementing standards-based accountability under No Child Left Behind: Responses of superintendents, principals, and teachers in three states*. Santa Monica, CA: RAND.

Stecher, B. M., S. Epstein, L. S. Hamilton, J. A. Marsh, A. Robyn, J. S. McCombs, J. L. Russell, and S. Naftel (forthcoming). *Pain and gain: Implementing No Child Left Behind in California, Georgia, and Pennsylvania, 2004 to 2006*. Santa Monica, CA: RAND.

APPENDIX A

Table A.1
Size of K-12 Public School Systems, 2005-2006

Feature	California	Georgia	Pennsylvania
Districts	1,128	204	730
Schools	9,863	2,489	3,250
Teachers	309,198	108,535	122,397
Students	6,437,202	1,598,461	1,830,684

SOURCE: NCES (undated).

Table A.2
Student Demographic Characteristics, 2003–2004

Characteristic	California (%)	Georgia (%)	Pennsylvania (%)
White	30	48	75
Hispanic	47	8	6
Black	8	38	16
Asian	11	3	2
Eligible for free/reduced-price lunch	48	50	31
ELLs	24	5	3
Students with disabilities	11	12	15

SOURCES: ELL data for Pennsylvania come from Pennsylvania Department of Education (2006). All other data come from NCES (undated).

Table A.3
District Sample and Cooperation, 2003–2004

Sampling	California	Georgia	Pennsylvania	Total
2003–2004				
Initial sample number	27	27	27	81
Replacement sample number	13	5	5	23
Total sample number	40	32	32	104
Number cooperating	19	25	24	68
Cooperation rate (%)	47.5	78.1	75	65.4
2004–2005 and 2005–2006				
Total sample number	56	37	39	132
Number cooperating	31	30	31	92
Cooperation rate (%)	55	81	80	70

Table A.4
School Sample and Cooperation, 2003-2004, 2004-2005, and 2005-2006

Sampling	California	Georgia	Pennsylvania	Total
2003-2004				
No. in sample	78	116	103	297
No. of cooperating schools	63	108	96	267
Cooperation rate (%)	80.7	93.1	93.2	89.9
2004-2005				
No. in sample	122	124	107	353
No. of cooperating schools	91	111	99	301
Cooperation rate (%)	75	90	93	85
2005-2006				
No. in sample	122	124	107	353
No. of cooperating schools	91	111	97	299
Cooperation rate (%)	74.6	89.5	90.7	84.7

Table A.5

Superintendent Survey Responses, 2003–2004, 2004–2005, and 2005–2006

Responding [Responses?]	California	Georgia	Pennsylvania	Total
2003–2004				
No. of cooperating districts	19	25	24	68
No. completing superintendent survey	18	20	22	60
Survey response rate (%)	94.7	80	91.6	88.2
2004–2005				
No. of cooperating districts	31	30	31	92
No. completing superintendent survey	24	24	19	67
Survey response rate (%)	77.4	80.0	61.3	72.8
2005–2006				
No. of cooperating districts	31	30	31	92
No. completing superintendent survey	26	21	20	67
Survey response rate (%)	83.9	70.0	64.5	72.8

Table A.6
Principal and Teacher Survey Responses, 2003–2004

Sampling and Responding	California	Georgia	Pennsylvania	Total
2003–2004				
No. of cooperating schools	63	108	96	267
No. of principal survey responses	51	88	88	227
Principal response rate (%)	80.9	81.5	91.7	85.3
No. in teacher sample	692	1,522	1,073	3,287
No. of teacher survey responses	487	1,318	926	2,731
Teacher response rate (%)	70.4	86.6	86.3	83.1
2004–2005				
No. of cooperating schools	91	111	99	301
No. of principal survey responses	78	95	87	260
Principal response rate (%)	86	86	88	86
No. in teacher sample	1,013	1,605	1,050	3,668
No. of teacher survey responses	826	1,409	938	3,173
Teacher response rate (%)	81.5	87.8	89.3	86.5
2005–2006				
No. of cooperating schools	91	111	97	299
No. of principal survey responses	75	96	88	259
Principal response rate (%)	82.4	86.5	90.7	86.6
No. in teacher sample	972	1574	1059	3605
No. of teacher survey responses	670	1366	931	2967
Teacher response rate (%)	68.9	86.8	87.9	82.3

APPENDIX B

Implementation of School-Based Accountability:
Salient Features of Sample Design and Sampling Weights

12/27/04

1. Three states in the study – California, Georgia and Pennsylvania. We will not combine the results across the three states.

2. Sampling Weights: The sampling weights are calculated separately for each state and at each of the three levels of the nested design: district, school, and teacher. At each level there are design weights associated with the probability that the unit was contacted. At the district and school level there are two response probabilities, the probability that the unit gives consent to participate in the study and the probability that the particular person completes their survey. At the teacher level there is only one response probability, the probability that they complete their survey. At the school and teacher levels there are two separate samples within each state – the elementary and middle school samples.

3. School districts in each state are divided into five strata, where each stratum is based on size of school district.

4. Strata size is based on the number of Elementary and Middle Schools, where strata definitions can be slightly different for each state.

Strata	CA	CA	GA & PA	PA	GA

	Strata Definitions	# Districts, # Contacted, #Cooperated, # Respond				Strata Definitions	# Districts, # Contacted, #Cooperated, # Respond				# Districts in Sample			
		N	Cnt	Cp	R		N	Cnt	Cop	R	N	Cnt	Cop	R
1	(#E = 0) OR (#M = 0 AND # E ≤ 2)	85	3	1	1	#E = 0 OR #M = 0	30	1	1	1	13	1	1	0
2	#E = 1 AND #M • 1	291	12	5	5	SAME	171	8	5	5	64	9	6	4
3	#E = 2 AND #M • 1	66	3	1	1	#E>1 AND #M = 1	220	11	6	6	38	6	4	4
4	#E • 3 AND #M ≤ 2	211	11	6	5	(#E=2 AND #M•2) OR (#E• 3 & #M=2)	40	5	5	5	21	4	3	2
5	#E • 3 AND #M • 3	158	12	6	6	SAME	37	7	7	5	41	13	11	10
Total		811	41	19	18		498	32	24	22	177	33	25	20

Issues with the sampling frame:
a) Some districts were omitted due to missing data on the # of elementary schools (PA 2 districts, GA 3 districts, CA 178 districts). According to JR Lockwood and Scott Naftel these schools were out of scope.
b) Three districts in GA were excluded because they were federally operated and thus out of scope. Two districts were selected into the original sample and then were found to have been taken over by the state (Oakland, CA and Philadelphia, PA). These were replaced by the first district on the randomized replacement list. Note that Philadelphia was originally selected with probability one.
c) Some small number of schools in each district had both elementary and middle grades. These schools were added to the count of both elementary schools and middle schools and were sampled as such (so the elementary grades, middle grades or both could be in the sample).

Districts were contacted in two stages:

All districts selected into the 'sample' were contacted (with the exception of (b) above). The districts selected as 'replacements' were given to WESTAT in a random order and they contacted needed replacements in order on this list. The number of replacement schools contacted is roughly equal to the number of districts in the sample that choose not to cooperate – although not exactly, presumably due to timing or lack of cooperation among contacted replacement schools.

5. Within strata, school districts are sampled with probability proportional to size, where the measure of size is the square root of the number of elementary schools in the school district. The number of elementary schools is a proxy for district size, and highly correlated with total number of schools in the district. If there were no elementary schools in the district the count was changed to one before taking the square root for the measure of 'size'. These dampened size variables were used to reduce the number of districts that were selected with probability close to one. There are three sets of weights at the district level.

   a. First is the probability of being contacted, based on the method SAS used to select the sample, this equals $n_h \dfrac{M_{hi}}{M_h}$ where n_h is the number of districts contacted in stratum h, M_{hi} is the 'size' of the individual district in stratum h ('size' is the square root of the number of elementary schools), and M_{hi} is the sum of the 'size' of all the districts in the stratum.
   If there is a district selected with probability one in the stratum, this district is considered its own strata. It has probability one of being contacted, its 'size' is not included in the stratum total, and the number of districts contacted in the strata is also reduced by one. The reciprocal of this is the weight associated with being contacted – in the code it is normalized to sum exactly to the total number of districts in the strata. This weight is used when comparing districts that cooperate with districts that do not cooperate.

   b. The next weight is associated with both being contacted and cooperating. It was noted that especially in CA, the districts that did not cooperate were more likely to have a higher percentage of schools not achieving AYP status. Since AYP status is likely to be correlated with outcomes of interest, we wanted to adjust the weights to more closely resemble the distribution of AYP status in each state. To achieve this goal three steps were taken. First, the 5 strata were collapsed into 2 strata where the first 4 strata categories became one stratum of 'smaller districts' and the 5th strata is 'larger districts'. This was done to reduce the sampling error in the estimated non-response probabilities while retaining the relevant attribute of the original strata. Second, districts were classified into high and low AYP status categories (based approximately on a median split in each state) and poststratification weights were constructed within the collapsed strata and AYP categories.

The weight is the number of districts in the state in the collapsed strata by AYP category cell divided by the number of districts in the cell that cooperated. These weights were then shrunk toward their mean to lessen the design effect created by adding a poststratification dimension, which then would be passed along to the lower levels of the sampling hierarchy (schools and teachers). The shrinkage lessened the degree to which the weighted sample recreated the AYP status distribution of the state but retained the collapsed stratification counts exactly. The amount of shrinkage applied varied by state with the shrinkage having a goal of reducing the design effect to approximately 1.3-1.4. In GA this did not require any shrinkage, in CA it required moderate shrinkage and in PA it required strong shrinkage. In PA the issue with AYP status was less of cooperation probability as it was of being contacted. By chance, the proportion of low AYP districts contacted was considerable lower than the state proportion. If there was a district selected with probability one its weight is not scaled up, only the weight of the remaining districts in the strata. These initial cooperation weights are combined with the contact weights to obtain the final cooperation weight. These weights are used to compare districts that respond to the survey to those that do not and get carried over to contribute to the school weights.

c. Finally, there are weights associated with being contacted, cooperating, and responding to the survey. These weights are the cooperation weights scaled up by the number who cooperate divided by the number who respond in the collapsed strata. This corresponds to a MAR assumption or MCAR within strata. Again, the weights for any districts selected with probability one are not scaled up. These weights are appropriate to use for analysis of the superintendent survey results.

d. Comments about district weights: It would have been preferable to adjust for AYP status at the school level because there are more sampled units and thus the design effect is likely to be smaller. However, the correlation between non-cooperation and AYP status was observed at the district level but not at the school level. So, we needed to adjust for it at the district level but did not want to pass on a huge design effect from the top level of the nested sampling to the levels below. This is why shrinkage was applied to the estimated non-cooperation weights but not to the contact weights since these were fixed by the design. Some of the shrinkage can be thought of as coming at no cost since it addresses sampling error in the estimation of cooperation probabilities within the strata/AYP status categories. However, this portion was estimated and in both CA and PA the level of shrinkage exceeds this portion. The remainder of the shrinkage is in the realm of bias-variance tradeoff. It would take considerable bias to overwhelm the variance created by the design effect if shrinkage was not used. (More specific estimates of this tradeoff are in the excel spreadsheet called DEFFcheck.)

6. Response rates across states were similar for GA and PA; with 24 and 25 school districts participating and a response rate about 75 percent of the schools contacted. In CA, the response rate was low (48 percent), with 19 school districts participating in the study.

7. Analyses do not require equal precision across strata (i.e., strata with smaller school districts have smaller samples than strata with larger school districts).

8. The sample of school districts drawn is disproportionately higher from the strata with the larger districts because there is a desire to include enough of the school districts with a larger share of the population of school-kids. (The sample should not include too many of the really small one or two school districts). In expectation, two Elementary and two Middle schools are invited to participate from each school district. Additional schools are recruited from larger districts to make up for the smaller districts, which cannot contribute two of each. The number of schools recruited from districts of various sizes is as follows:

Elementary Schools

CA		PA		GA	
#schools	# contacted	# schools	#contacted	# schools	# contacted
1	1	1	1	1	1
2-5	2	2-4	2	2-9	2
6-10	3	5-7	3	10-14	3
11-19	4	8-15	4	15+	4
20+	5	16+	5		

Middle Schools

CA		PA		GA	
#schools	# contacted	# schools	#contacted	# schools	# contacted
1	1	1	1	1	1
2	2	2	2	2	2

3	3	3	3	3-4	3
4	4	4	4	5+	4
5+	5	5+	5		

There are three sets of weights created at the school level for each of the elementary, middle school and combined samples in each state, equivalent to the weights created at the district level: contact, cooperation, and response weights. In addition there are alternate weights calculated for the cooperation and response weights.

   i.   Within any district that cooperated, the probability that a school is contacted depends on the number of schools in the district and is given in the tables above. The initial contact weight is the number of schools (either elementary or middle) in the district divided by the number contacted. These weights are fixed by the design. Final school contact weights are obtained by multiplying the district cooperation weight by the school contact weight. These weights are used to test for differences between cooperating and non-cooperating schools.

  ii.   Cooperation rates varied by state and school level (elementary/middle). For the elementary and middle school samples respectively the rates in each state are: CA, 78 percent and 84 percent, GA, 88 percent and 97 percent, and PA, both 90 percent. We assume that cooperation is random within district (MAR or MCAR within district) and so initial cooperation weights are the number of schools contacted divided by the number of schools cooperating in each district. In each state there are some districts that are dropped due to non-cooperation at the school level. The district level cooperation weights are adjusted for this non-response by scaling up the weights within the collapsed strata to account for the district(s) lost due to school non-cooperation. The initial cooperation weights are multiplied by the contact weights to obtain the final cooperation weights. These weights are used to compare responding and non-responding schools and are passed on to create the teacher level weights.

 iii.   Survey response weights are created equivalently to cooperation weights. For the elementary and middle school samples respectively the rates in each state are: CA, 77 percent and 84 percent, GA, 88 percent and 97 percent, and PA 93 percent and 91 percent. We assume that survey response is random within district (MAR or MCAR within district) and so initial response weights are the number of cooperating schools divided by the number of principals completing their surveys in each district. In each state there are some districts that are dropped due to non-cooperation and/or non-response at the school level. The district level cooperation weights are adjusted

for this non-response by scaling up the weights within the collapsed strata to account for the district(s) lost due to school non-cooperation and non-response. The initial response weights are multiplied by the final cooperation weights to obtain the final response weights. These weights are used to analyze the principal survey results.

iv.   It was unclear at the time that the sampling weights were created whether the elementary and middle school samples would ever be combined for analysis. To prepare for this possibility, weights were created for the combined samples within each state. There is a minor issue in creating these weights because they have slightly overlapping frames – schools that contain both elementary and middle grades are in both frames and could be come into the combined sample from either or both frames. The probability of being contacted for these schools is half their probability of being contacted in a particular sample if they are only in one and the average of their probabilities of being contacted in both samples if they are selected to be in both (see Kish, 1995 for reference). In addition, the weights of all schools in the sample are scaled down slightly to account for the difference between the sum of the number of schools in the overlapping sample frames and the population size. Except for this issue, the three sets of weights for the combined sample are created as described above.

v.   The cooperation and response weights were created within district to maintain the nested structure of the sampling design in anticipation of using hierarchical models to analyze the data.

vi.   Alternate cooperation and nonresponse weights were constructed that do not retain the nested structure of the sampling design (the weights of the schools within a district do not add to the number of schools within that district). These non-nested weights were constructed because the nested weights above were created a distribution of school covariates that was far from the actual distribution for CA elementary schools. In CA at the elementary school level, districts with a high proportion of Title I schools are less likely to be in the sample. In addition, Title I schools were less likely to cooperate and the principals were less likely to complete the surveys than non-Title I schools. The result is that the original weights underestimated the proportion of Title I elementary schools in the state. To address this, alternate cooperation and non-response weights were calculated in each state at each schooling level. These weights post-stratified to the number of schools in the two District level collapsed strata by Title I school status. This helped considerably in CA elementary schools and tended to make minor changes in other states and levels where the weighted estimates of the Title I proportion were not far off with the original nested weights. Probably by chance, the design effects using the alternate weights tended to decrease for elementary schools and increase for middle schools. These weights are used for non-hierarchical analyses.

9. All math and science teachers within cooperating schools are contacted to complete a survey. Conditional on school cooperation, every teacher's probability of being contacted is 1. The teacher survey non-response rates for elementary and middles school samples respectively in each state are: CA, 72 percent and 70 percent, GA, 84 percent and 85 percent, and PA, 87 percent and 82 percent. Initial teacher survey response rates were created by dividing the number of teachers contacted by the number of teachers completing their survey within school. This assumes that survey responses are missing at random within school (MAR or MCAR within school). Teacher survey response was still correlated with some school attributes once district attributes were accounted and so teacher non-response rates were estimated within school not district. This difference only slightly increased the design effect in most cases. Estimating non-response weights within school maintains the nesting of the sample. (In PA there was one middle school with very low teacher response rates in a large district. These non-response weights were very high outliers but the school was not from an uncommon stratum or AYP status category. If left alone these weights alone inflate the DEFF enormously (from 2 to 3). The probability that this increase in variance is worth the decrease in bias is extremely unlikely. These initial weights were trimmed back to the maximum of the initial middle school non-response weights.) The initial weights were multiplied by the final school cooperation weights to obtain final teacher survey weights. These weights are used to analyze the teacher survey results. There are separate weights for the elementary, middle, and combined samples.

    a. The alternate weights (see 8. vi. above) at the school level were also used to create alternate weights at the teacher level. The initial teacher weights were created as described above but combined with the alternate non-nested school level weights instead of the nested original weights. These weights are used for non-hierarchical analyses.