

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection methods to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection has been conducted previously, include the actual response rate achieved during the last collection.

This survey will target six distinct populations of interest (superintendents, principals, teachers, higher education administrators, PTA presidents, state and local Chamber of Commerce presidents), each of which is operationally defined in Table 1.B below. Each of the populations will be sampled separately in each of the five states in the REL Southwest region (Arkansas, Louisiana, New Mexico, Oklahoma, Texas), and in some cases separately for rural and non-rural school districts.

Table 1.B Operational Definitions of each Population of Interest (Universe)

Population	Operational Definition
Superintendents	All superintendents of a public K-12 school district in the five state region
Principals	All principals of a public K-12 school in the five state region
Teachers	All regular (not substitute) classroom teachers in a public K-12 school in the five state region
Higher Education Administrators	All Admissions Directors and Student Affairs Directors of public and private 2 and 4 year colleges and universities in the five state region
PTA Presidents	All presidents of a PTA for a public K-12 school in the five state region
Chamber of Commerce Presidents	All presidents of a state or local Chamber of Commerce in the five state region

There will be no attempt to generalize any of the survey results beyond the populations as operationally defined in Table 1.B. It would not be feasible to collect a random sample of all parents in the region, therefore PTA presidents were selected because they are actively involved in the schools, have a good understanding of the concerns of parents in their school, and are a population that can be fully enumerated and randomly sampled. They are, of course, only a select subsample of the broader parent population, and results of the survey will only be generalized to all PTA presidents, and not all parents.

Similarly, while state and local Chamber of Commerce presidents are only a subpopulation of all of the business leaders in the five state region, they are an important subpopulation that can be feasibly defined and sampled. Results of the sample will only be generalized to the full subpopulation of all state and local Chamber of Commerce presidents, and not the full population of all business leaders.

Table 2.B below provides estimates of the various universe sizes. This data is based primarily on estimates from the Market Data Retrieval (MDR) 2005-2006 Educator Database, a prominent national level educator data collector. In addition, estimates from Why Rural Matters 2005 (Johnson & Strange, 2005) are used to modify the MDR data to estimate the universe sizes for rural and non-rural subpopulations. Estimates for state and local Chamber of Commerce presidents come from a comprehensive online directory of U.S. Chamber of Commerce organizations (www.2chamber.com).

Table 2.B Estimated Universe Sizes

Population	AR	LA	NM	OK	TX	Total
Superintendents	270	76	99	561	1,072	2,078
Rural only	132	24	35	286	247	724
Non-rural only	138	52	64	275	825	1,354
Principals	1,046	1,418	769	1,661	7,398	12,292
Rural only	513	454	269	847	1,702	3,785
Non-rural only	533	964	500	814	5,696	8,507
Teachers	35,168	52,938	19,142	43,998	328,666	479,912
Rural only	12,344	13,605	3,503	13,947	42,069	85,468
Non-rural only	22,824	39,333	15,639	30,051	286,597	394,444
Higher Education Administrators	79	116	56	94	375	720
PTA Presidents	1,046	1,418	769	1,661	7,398	12,292
Rural only	513	454	269	847	1,702	3,785
Non-rural only	533	964	500	814	5,696	8,507
Chamber of Commerce Presidents	138	110	66	167	584	1,065

The data in Table 2.B is based on MDR data which provides reasonably good estimates of the universe sizes, but which will not lead to a sufficiently complete sampling frame for each population of interest. Therefore, in most cases, the sample of respondents will be obtained using multi-stage sampling which will provide us with a sample that is equivalent to one obtained with simple random sampling from a complete frame, if such a frame were feasible to construct. Details of the sampling process for each of the populations are provided below as part of the answer to question #2.

The response rate for the survey is expected to be at least 90%. Details about how respondents will be defined, and efforts that will be employed to maximize response rate are detailed in the answer to question #3 below. This is the first time that this survey is being conducted, so no previous response rates are available.

2. Describe procedures for the collection of information including: statistical methodology for stratification and sample selection; estimation procedure; degree of accuracy needed for the purpose described in the justification; unusual problems requiring specialized sampling procedures; and any use of periodic (less frequent than annual) data collection cycles to reduce burden.

We are collecting data for six different populations for this survey (superintendents, principals, teachers, higher education administrators, PTA presidents, state and local Chamber of Commerce presidents) across five different states (Arkansas, Louisiana, New Mexico, Oklahoma, Texas) and, in some cases, two levels of urbanicity (rural, non-rural). Operational definitions of the various populations of interest are provided in Table 1.B above.. Data for the different populations will never be combined in any way, although within a population, data will be aggregated across both state and urbanicity.

Although the sampling methodology is quite similar in most cases, there are certain differences, and so for clarity sake, we will outline the sampling approach for each population separately. In

almost all cases, the identification of a complete sampling frame is not feasible, and so some variation of multi-stage sampling is used to obtain a sample identical to one that would have been obtained from simple random sampling for a complete frame. As appropriate, the selection of clusters will always be with probability proportional to estimated size (PPES), and the estimated size will of course vary depending on the population being sampled, and stage of the sampling.

Superintendents

In the case of superintendents, it will actually be possible to specify complete sampling frames. It will be possible within each state to obtain a complete list of public K-12 school districts, and using the definition as laid out in *Why Rural Matters 2005* (Johnson & Strange, 2005), the list will be subdivided into two complete lists of the districts: rural public K-12 school districts for each state, and non-rural public K-12 school districts for each state.

Because there is only one superintendent per district, simple random sampling from the list of school districts is equivalent to simple random sampling of superintendents themselves.

The sample of superintendents will be stratified by both urbanicity and by state; although an equal sample size will be collected in each of these strata. When the data is combined across cells, it will be done so using weights reflecting the universe sizes corresponding with the various cells.

Principals

It will not be possible to obtain a complete frame for principals, so a multi-stage sample will be collected instead. For each state and urbanicity level, a complete list of school districts will be obtained. A school district will be randomly selected from this list with probability proportional to estimated size. Once a school district is selected, then a complete list of schools in that district will be obtained, and one of those schools will be randomly selected (this time with equal probability since there is one principal per school). This process will be repeated a number of times to obtain the desired sample size (as shown in Table 3.B below).

The sample of principals will be stratified by both urbanicity and by state; although an equal sample size will be collected in each of these strata. When the data is combined across cells, it will be done so using weights reflecting the universe sizes corresponding with the various cells.

Teachers

It will not be possible to obtain a complete frame for K-12 public school teachers, and so a multi-stage sample will be collected instead. For each state and urbanicity level, a complete list of school districts will be obtained. A school district will be randomly selected from this list with probability proportional to estimated size. Once a school district is selected, then a complete list of schools in that district will be obtained, and one of those schools will be randomly selected with probability proportional to estimated size. From this school, 4 teachers will be obtained at random. This process will be repeated a number of times to obtain the desired sample size (as shown in Table 3.B below).

The sample of teachers will be stratified by both urbanicity and by state; although an equal sample size will be collected in each of these strata. When the data is combined across cells, it will be done so using weights reflecting the universe sizes corresponding with the various cells. In general, the schools sampled for teachers will not be the same as those sampled for principals, since the sampling probabilities for the school will vary relative to the two populations of interest.

Higher Education Administrators

A complete frame of all College Admissions Directors and Student Affairs Directors for both public and private colleges and universities will be obtainable from MDR, and this list will be used within each state to select a simple random sample (or a census in some states where this is feasible).

The sample of higher education administrators will be stratified by state only, with an equal sample size being collected in each state (except where a census is appropriate due to small universe size). When the data is combined across states, it will be done so using weights reflecting the appropriate universe sizes.

PTA Presidents

For purposes of this survey, we are looking only at the subpopulation of PTA presidents, and only generalizing the results of the survey to this subpopulation. The sampling process will be similar to that of principals, since there is for each school, one PTA president, just like there is one principal. In the case that a particular school had no PTA, that would simply be a school not in the defined universe, and a different school and corresponding PTA president would be sampled instead. In general, it will be most efficient to co-sample principals and PTA presidents from the same sample of schools.

The sample of PTA presidents will be stratified by both urbanicity and by state; although an equal sample size will be collected in each of these strata. When the data is combined across cells, it will be done so using weights reflecting the universe sizes corresponding with the various cells.

Chamber of Commerce Presidents

For purposes of this survey, we are looking only at the subpopulation of state and local Chamber of Commerce presidents, for those listed in each town as a business association or chamber. A complete list of such organizations will be developed for each state. Simple random sampling will then be used to develop the sample to be selected.

The sample of state and local Chamber of Commerce presidents will be stratified by state only, with an equal sample size being collected in each state. When the data is combined across states, it will be done so using weights reflecting the appropriate universe sizes.

Target Sample Sizes

Table 3.B below provides the sample sizes for each cell that will be needed to obtain the desired level of precision assuming a 90% response rate.

Table 3.B Target sample sizes per cell for desired precision assuming 90% response rate.

Population	Urbanicity	AR	LA	NM	OK	TX	Total
Superintendents	Rural only	40	40	35**	40	40	195
Superintendents	Non-rural only	40	40	40	40	40	200
Principals	Rural only	40	40	40	40	40	200
Principals	Non-rural only	40	40	40	40	40	200
Teachers	Rural only	64*	64*	64*	64*	64*	320
Teachers	Non-rural only	64*	64*	64*	64*	64*	320
Higher Education Administrators	All	79**	89	56**	89	89	402
PTA Presidents	Rural only	40	40	40	40	40	200
PTA Presidents	Non-rural only	40	40	40	40	40	200
Chamber Presidents	All	89	89	66**	89	89	422
Total	All	536	546	485	546	546	2659

* Teachers are sampled at 16 sites with 4 per site

** These sample sizes are smaller than in other states because they represent the entire population of interest

The primary results of interest in this survey will be estimates of the population proportions agreeing with various items. Both point estimates and corresponding 95% confidence intervals will be provided. Estimates will be provided separately for each population and at various levels of aggregation. The most important results are for the highest level of aggregation (results for the entire region), and it is desired that this have precision level of at least ± 0.10 at the 95% confidence level. Results for lesser levels of aggregation (e.g., within a single state, or across states for rural only) are desired to have a precision level of at least ± 0.15 at the 95% confidence level.

As can be seen in Tables 3.B and 4.B, the targeted sample sizes provide the desired level of precision. We'd like to emphasize a number of points about these results.

The error bands provided in Table 4.B are estimates of the worst case scenario (at $p=.50$), which is the conservatively appropriate estimate, but as p varies from $.50$, the confidence intervals will of course be correspondingly smaller.

For teachers, since 4 teachers are chosen at each school during the multi-stage sampling process, the effective sample size (ESS) must be calculated. The ESS provides the sample size from a simple random sample that has the equivalent precision the actual sample size from the complex design. For multi-stage sampling, this is given by

$$ESS = n / Deff$$

where n is the actual sample size and $deff$ is the design effect.

The design effect is calculated by

$$Deff = 1 + ICC(k - 1)$$

where ICC is the Intraclass correlation coefficient and k is the number of subjects sampled within a cluster.

Estimation of these values prior to data collection for purposes of sample size determination must be based on values from similar surveys previously conducted. The most similar survey we are aware of is the Schools and Staffing Survey conducted by the National Center for Education

Statistics (NCES). Communications with NCES staff led to the suggestion of using a technical report called Design Effects and Generalized Variance Functions for the 1990-91 Schools and Staffing Survey (SASS). This document does report various design effects, but unfortunately neither ICCs nor cluster sizes, although maximum cluster sizes were reported. The design effects for the items most similar to our survey were approximately 2, and using different estimates of cluster sizes, we estimate the corresponding ICC to range about .05 to .20. Erring on the conservative side, we selected an ICC of .20. For the cluster size of 4 teachers per school, this yielded a value of Deff = 1.6. In order to have the same effective sample size as the other populations in which multiple subjects were not taken from a cluster, it will be necessary to sample 16 schools with 4 teachers each, making the assumption of 90% response rate.

Since the samples are being collected within a state (and sometimes within urbanicity strata within a state), aggregation of these samples will be necessary to obtain the point estimates for the larger populations of interest. This will be done by calculating the appropriate weighted proportion, and corresponding confidence intervals. Again, this will require the calculation of an effective sample size, although in this case the formula for ESS is different. For a weighted proportion, this is given by

$$ESS = \frac{(\sum_{i=1}^n w_i)^2}{\sum_{i=1}^n w_i^2}$$

Where w_i is the weight assigned to the i th person in the sample, based upon the strata that person belongs to.

Finally, it should be noted that for certain cells in the design (e.g., rural superintendents, local and state chamber of commerce presidents in New Mexico), the population as operationally defined is relatively small. In some of these cases we will actually be collecting a census of the population as defined, and in other cases, a very large percent of the finite population will be collected. The values in Table 4.B do not include a finite correction factor, and therefore for those particular populations, the precision is underestimated in Table 4.B.

Table 4.B Estimated Minimum Precision for 95% Confidence Intervals

Subpopulation	95% CI ±
Superintendents	
5 state region	.09

5 state region (rural only)	.10
5 state region (non-rural only)	.12
Arkansas	.12
Louisiana	.13
New Mexico	.13
Oklahoma	.12
Texas	.14
Principals	
5 state region	.09
5 state region (rural only)	.10
5 state region (non-rural only)	.12
Arkansas	.12
Louisiana	.13
New Mexico	.13
Oklahoma	.12
Texas	.14
Teachers	
5 state region	.09
5 state region (rural only)	.09
5 state region (non-rural only)	.12
Arkansas	.13
Louisiana	.13
New Mexico	.14
Oklahoma	.13
Texas	.15
Higher Education Administrators	
5 state region	.08
Arkansas	.12
Louisiana	.11
New Mexico	.14
Oklahoma	.11
Texas	.11
PTA Presidents	
5 state region	.09
5 state region (rural only)	.09
5 state region (non-rural only)	.12
Arkansas	.13
Louisiana	.13
New Mexico	.14
Oklahoma	.13
Texas	.15
Chamber of Commerce Presidents (local & state)	
5 state region	.08
Arkansas	.11
Louisiana	.11
New Mexico	.11
Oklahoma	.11
Texas	.11

3. Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided

for any collection that will not yield “reliable” data that can be generalized to the universe studied.

Once a sample has been identified for each population of interest, e-mail addresses and phone numbers will be obtained for each person in the sample. Each person in the sample will first be sent a postcard alerting them to the upcoming survey, and will then be contacted by email, and asked to participate in the on-line survey. Each person will be sent no more than 4 follow-up emails asking them to participate. If necessary, each person who has not yet taken the survey will be contacted then by phone. Up to 5 attempts will be made to reach a person, but no more than one phone contact will be made with any person (unless they agree to participate in the phone survey at a later time, in which case a second phone call at an agreed upon time will be made). Anyone who does not respond to the survey after these contacts will be considered a non-respondent.

With respect to maximizing the response rate of the various K-12 samples, we will also be sending an informational letter to superintendents if either they or any of their principals will be included in the survey, and will send an informational letter to principals if either they or any of their teachers will be included in the survey. The Superintendent letter will provide the names of all persons in his/her district being sampled (superintendent, principals, teachers, PTA presidents as appropriate). The Principal letter will provide the names of all persons in his/her school being sampled (principal, teachers, PTA president as appropriate). We will request that the Superintendent or Principal encourage the identified persons to complete their individual surveys as their opinions are important. The encouragement of Superintendents and Principals should assist in maximizing the response rate of the surveys.

It is anticipated that there will be at least a 90% unit response rate, and considerable effort will be made to maximize response rates:

- Email invitation will be kept short and will provide only one link to the survey.
- Email invitation will identify the Regional Educational Laboratory - Southwest and briefly define:
 - o Purpose of study
 - o Length of survey
 - o Privacy statement
 - o Explanation of how we will use responses
 - o Clarity of short / concise instructions, purpose and questions
- Respondents will be informed that research is being conducted to aid the Department of Education in fulfilling its education research objectives through the Institute of Education Sciences' Regional Educational Laboratory - Southwest.
- Respondents will also be informed that participation in the survey is voluntary.
- All respondents will be informed about the confidential nature of this survey and data collection and will be assured of the usage of this information for research purposes only.
- Survey will include a limited number of questions to reduce time respondents will need to complete.
- Majority of questions will be closed/multiple choice with limited number of open-ended questions.

Should the unit response rate drop below 80% for any population, a non-response bias analysis will then be conducted. Given the fairly small sample size likely for non-respondents in a particular subpopulation, it will most likely be feasible to attempt to contact non-respondents and provide them with extra incentives to participate and yield data for the non-response bias analysis. If such data cannot be collected, then auxiliary demographic data will be used, since this information should be available for both respondents and non-respondents. This would include publicly available information such as length of time in position, education level, etc. Again, it is not anticipated that unit non-response will actually be a problem.

Given the nature of the survey, item non-response is even less likely to be a problem. However, should missing data be a problem, then multiple imputation will be used to address the item non-response.

4. Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of data.

The pilot is tentatively scheduled for June 4, 2007, dependent on OMB approval of this submission.

No pilot or data collection with more than 9 people will occur until OMB project approval has been granted. Therefore, pilot date is subject to change. (Please see pilot and data collection schedule on page 13 of Supporting Statement, Part A.)

A pilot of the survey will be conducted prior to launching the survey to the entire sample. In this pilot, data will be collected from less than 9 respondents and results will be reviewed to “minimize burden and improve utility” by testing the average time it takes to complete the survey using the online version and testing among a range of titles, to ensure that questions are clear and understandable to the person taking the survey; and to help ensure that the online survey taking process runs smoothly and without technical problems.

5. Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

Statistical Consultant: Brian A. Harris-Kojetin, Chief Methodologist, Office of Statistical Policy, Office of Management and Budgeting
Phone: 202-395-7314

Statistical Consultant: Ralph Lee, Mathematical Statistician, National Center for Education Statistics
Phone: 202-502-7381

Vendor name: Galloway Research
Vendor contact: John Ucab
Title: Director, Telephone Survey Department
Role: Responsible for monitoring client phone survey and ensuring that data collection is done according to client guidelines.
Phone: 210-734-4346

Vendor name: Market Data Retrieval
Vendor contact: Beth Shields
Title: Account Manager, Market Data Retrieval
Role: Responsible for providing contact information of educators, for data collection.
Phone: 312-345-4355