

America's Challenge: Effective Teachers for At-Risk Schools and Students

Chapter 6 Methodology

Getting Started: A Survey of New Public School Teachers on Their Training and First Months on the Job

This survey includes interviews with a nationally representative sample of 641 first-year school teachers throughout the continental United States. We also conducted oversamples of teachers who participated in alternative teaching certification programs. Those interviews are not included in this analysis but will be included in subsequent reports on this data. The survey was conducted by Princeton Survey Research Associates International (PSRAI). Data were collected by telephone (324 interviews) by Princeton Data Source, LLC, and online (324 surveys) by PSRAI between March 12, 2007, and April 23, 2007.

In designing the survey questions and sample, Public Agenda conducted interviews with leading experts both from traditional schools of education and from alternative programs to discuss the sampling frame and the topics to explore in the survey. The National Comprehensive Center for Teacher Quality (NCCTQ) and the Farkas-Duffett Research Group (FDR) were consulted further regarding sampling, survey topics, and questionnaire design. Prior to interviewing, the questionnaire was tested multiple times with two new teachers and items were recrafted based on these pretests. In addition, random interviews were monitored to insure the quality of the interviews and that the questions were clear and answerable.

The sample includes oversamples of teachers in both Midwest and high-needs schools. The final data were weighted to account for the disproportionate sample design. Final results based on the general sample are representative of all first-year teachers' continental U.S. public schools. The margin of sampling error for the complete set of weighted data is ± 4 percent. Details on the design, execution, and analysis of the survey are discussed below.

Questionnaire Design

Respondents were asked 111 items. These included screener questions to ensure that respondents were first-year teachers, demographic questions to describe the teachers who took part in our survey, and closed-ended opinion questions. The questionnaire uses a blend of different kinds of questions, some of which tackle similar issues in different ways. Most questions ask respondents to use a scale (either three or four points) to rate different aspects of their training or experiences teaching and to measure the strength of various beliefs they may have about teaching.

Many of our four-point scales are Likert scales, in which we ask the degree to which a respondent accepts a particular statement. In the report, we often collapse the choices to the nominal level by combining the positive and negative responses. (Collapsing Likert scales into their nominal components [agree/disagree] is a commonly used technique in public opinion

research. After transforming the data, it is subject to chi-square assessments.) Those interested in seeing the degree to which someone agreed or disagreed to the statement can consult the charts in the reports, which break out strength of acceptance.

We also used questions in which respondents were asked to choose between two mutually exclusive and balanced statements involving tradeoffs. Analyzed in context with other results, these “forced-choice” items shed light on respondents’ priorities and avoid the central tendency bias inherent in Likert-style questions. The choices themselves may be artificial, but they typically echo natural language gleaned from qualitative research. This questionnaire reflects the language and expressions used by teachers during focus groups for this project and from previous research with teachers.

For example, one of the questions asked new teachers, “Which comes closer to your view?”

1. I may be new to teaching, but compared to what other teachers are doing, my students are probably lucky to have me. [OR]
2. I’m sometimes afraid that my students are paying a heavy price because of my lack of experience.

This item is drawn directly from the qualitative research where a new teacher said in a focus group, “I’m a teacher to these kids. I’m not qualified at all. Yet, I’m still possibly better than what could be there. It’s absolutely ridiculous.” Because the other teachers in the focus group agreed with this perspective, we decided to counter-balance the notion that students are lucky to have a new teacher with one that gives an equally reasonable, but very different response. In this instance, the presentation of the alternative viewpoint is intended to test and probe whether this response is strongly held even when positioned against a robust alternative.

In a few instances, the questionnaire contains compound questions combining two seemingly separate concepts. The decision to combine concepts within a single item mirrors the way teachers discuss and couple ideas in focus groups.

For example, one item in our series of questions about potential drawbacks to teaching is, “There is so much testing and not enough freedom to be creative.” This item mirrors a comment by a new teacher in a focus group who said, “I think it’s absolutely a matter of testing taking away too much time... You are very restricted in the amount of time that you have to try new, creative theories, because you have to get this, this, and this in before.”

Obviously, compound items could be asked separately, and other researchers may wish to tease them apart based on the results here; indeed, we invite them to do so. However, we believe these compound items capture authentic and useful information about new teachers’ overall priorities and concerns and are consistent with previous studies conducted by Public Agenda.

Sample and Data Collection Procedures

Sample

The sample was drawn from Market Data Retrieval's (MDR) New Teachers list. MDR (www.marketdataretrieval.com/mdredlists.asp), a company of Dunn & Bradstreet, is a leading provider of marketing information and services for the education market. The list was stratified by region and high-needs status so that schools in the Midwest and schools considered high-needs could be oversampled. The Midwest was defined as seven states: Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin. Schools were considered high-needs when at least 51 percent of students received free or reduced-price lunch. For sampling, schools were considered high-needs if a majority of the students were eligible for a free or reduced-price lunch program as reported in the MDR sampling frame (i.e., LUNCHPRGM/EXACTERL >= 0.51). For analysis purposes, teachers were identified as working in high-needs schools if they answered that a majority of their school's students were eligible for a free or reduced-price lunch program (D3 = 3). This variable is similar to the high-needs variable used in the sampling phase. Prior to starting each interview, respondents were screened to confirm that they were first-year classroom teachers.

Contact Procedures

Data for the sample were collected by telephone and online between March 12, 2007, and April 23, 2007. Teachers in the general sample were mailed advance letters explaining the purpose of the study and encouraging their participation. The advance letters included a web address and password so respondents could complete the interview online. A follow-up e-mail also was sent by MDR to all teachers in the sample who had a valid e-mail address. Daytime phone interviews were conducted by interviewers at Princeton Data Source (PDS).

Two batches of general MDR samples were sent out during the field period. An initial wave of 3,592 first-year teachers (2,688 not high-needs and 904 high-needs) from the general MDR sample first were contacted through advance letters mailed on Thursday, March 8. On Wednesday, March 14, PDS began daytime calling of teachers in this first batch who had not yet completed the interview online. E-mail reminders for the first batch were sent out by MDR on Friday, March 23, to those who had not yet completed the interview by phone or online.

A second wave of 2,504 first-year teachers (673 not high-needs and 1,831 high-needs) from the general MDR sample were first contacted through advance letters mailed on Monday, April 2. On Monday, April 9, PDS began daytime calling of teachers in this second batch who had not yet completed the interview online. E-mail reminders for this second batch were sent out by MDR on Thursday, April 19, to those who had not completed the interview by phone or online. Samples and quotas were tracked daily so that the telephone and online data collection modes remained coordinated. For the general sample, 324 interviews were completed by phone and 317 surveys were completed online.

Interviewers with experience in reaching and gaining cooperation from hard-to-reach populations were used for this project. All interviewers received training on this questionnaire and were

briefed on special refusal aversion techniques that apply to this population. Ten percent of all interviews were monitored by supervisors who could identify problems and challenges and coach interviewers. In addition, Public Agenda staff reviewed recordings of three interviews recorded on the first full day of interviewing. Interviewers accommodated respondents' schedules and arranged appointments. Respondents also were given a toll-free phone number to call to schedule an appointment or complete an interview at their convenience.

Weighting and Analysis

The data were weighted to account for the oversampling of key groups. Population parameters came from two sources. The regional distribution parameter was defined as the regional distribution of all teachers for 2003 as reported by the National Center for Education Statistics (NCES). The high-needs parameter was defined by first computing percentage of teachers teaching in at risk schools in each state individually. This within-state, at-risk percentage was taken from the MDR new teacher sampling frame. Then the states were combined in the proportions used in the regional distribution. The final estimated population distribution used from weighting is given in Table 1.

Table 1. Estimated Population Distribution

	High Needs	Not High Needs	Total
Midwest	3.3%	14.9%	18.2%
Not Midwest	31.3%	50.5%	81.8%
Total	34.6%	65.4%	100.0%

The first step in the weighting was to correct for the disproportionate sampling of the general teacher list. The weighting ensures that the final general sample region and high-needs percentages match the population distribution. Table 2 compares unweighted and weighted general sample demographics to population parameters.

Table 2. Traditional Sample Weighting

	Population Parameter	Unweighted Sample	Weighted Sample
High-Needs			
Midwest	3.3%	5.9%	3.3%
Not Midwest	31.3%	33.1%	31.4%
Not High-Needs			
Midwest	14.9%	17.2%	14.8%
Not Midwest	50.5%	43.8%	50.5%

Effects of Sample Design on Statistical Inference

Postdata collection statistical adjustments require analysis procedures that reflect departures from simple random sampling. PSRAI calculates the effects of these design features so that an

appropriate adjustment can be incorporated into tests of statistical significance when using these data. The so-called “design effect,” or *deff*, represents the loss in statistical efficiency that results from disproportionate sampling and systematic nonresponse. The total sample design effect is 1.27.

PSRAI calculates the composite design effect for a sample of size *n*, with each case having a weight, *w_i*, as:

$$deff = \frac{n \sum_{i=1}^n w_i^2}{\left(\sum_{i=1}^n w_i \right)^2} \quad (1)$$

In a wide range of situations, the adjusted standard error of a statistic should be calculated by multiplying the usual formula by the square root of the design effect (\sqrt{deff}). Thus, the formula for computing the 95 percent confidence interval around a percentage is:

$$\hat{p} \pm \left(\sqrt{deff} \times 1.96 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right) \quad (2)$$

where \hat{p} is the sample estimate and *n* is the unweighted number of sample cases in the group being considered.

The survey’s margin of error is the largest 95 percent confidence interval for any estimated proportion based on the total sample—the one around 50 percent. For example, the margin of error for the entire sample is ± 4 percent. This means that in 95 out every 100 samples drawn using the same methodology, the estimated proportions based on the entire sample will be no more than 3.8 percentage points away from their true values in the population. It is important to remember that sampling fluctuations are only one possible source of error in a survey estimate. Other sources, such as respondent selection bias, questionnaire wording, and reporting inaccuracy, may contribute additional error of greater or lesser magnitude. Table 3 shows design effects and margins of error for key sample subgroups.

Table 3. Design Effects and Margins of Error for Key Subgroups

	<i>n</i>	Design Effect	Margin of Error
Total sample	641	1.09	4.0%
High-needs	315	1.10	5.8%
Not high-needs	326	1.07	5.6%

Subgroup Analysis

Although most of the findings in this report are derived from the total sample, additional significance tests were employed when comparing survey results across subgroups. Any differences reported between groups were deemed significant after considering the overall margin of error, sampling tolerance, and Pearson’s chi-square calculations for each question.

Response Rate

Table 4 reports the sample disposition for the sample. The response rate estimates the fraction of all eligible respondents in the sample that ultimately were interviewed. At PSRAI it is calculated by taking the product of three component rates (PSRAI’s disposition codes and reporting are consistent with the American Association for Public Opinion Research standards):

- Contact rate—the proportion of working numbers where a request for interview was made—of 32 percent
- Cooperation rate—the proportion of contacted numbers where a consent for interview was at least initially obtained versus those refused—of 89 percent
- Completion rate—the proportion of initially cooperating and eligible interviews that were completed—of 99 percent

Thus, the response rate for this survey was roughly 29 percent.

Table 4. General Sample Disposition

Size	Description
5868	T Total Pieces of Sample
65	OF Not a working phone number
25	OF Computer/Fax
5778	Working numbers
98.5%	Working Rate
715	UH No Answer
94	UH Busy
1356	UO_{NC} Answering Machine
1735	UO_R Callbacks
0	NC Non-Contacts after determined eligible
20	UO_{NC} Other Non-Contacts
1858	Contacted numbers
32.2%	Contact Rate

Size	Description
202	UO_R Refusal 1 - Refusal before eligibility status known
0	R Refusal 2 - Refusal after case determined eligible
1656	Cooperating numbers
89.1%	Cooperation Rate
331	IN1 Teacher no longer with school
681	IN2 Not a first year teacher
644	Eligible numbers
38.9%	Eligibility Rate
3	R Interrupted
641	I Completes
99.5%	Completion Rate
28.5%	Response Rate

The Focus Groups

Focus groups allow for an in-depth, qualitative exploration of the dynamics underlying the public’s attitudes toward complex issues. Insights from participants in these focus groups were important to the survey design and actual quotes were drawn from the focus groups to give voice to attitudes captured statistically through the surveys. All focus groups were moderated by Public Agenda senior staff.

Four focus groups were conducted. One was conducted with participants in an alternative certification program in the Philadelphia region. Two more also were held in Philadelphia, one with senior education majors and master’s-plus students from an urban university and one with the same population from a suburban university. The last group was done in Chicago with first-year teachers in an urban alternative certification program and with urban master’s-plus students.