

## Technical Appendix

This technical appendix provides detail on the research design and methodological procedures used in the national survey of parents and teachers. We begin by outlining the procedures employed in designing the two questionnaires. This is followed by a detailed description of the sampling protocols, first for the parent survey and then for the teacher survey. Following this is an assessment of the quality of the fieldwork both with respect to response rates and sample coverage. Then we discuss sample weighting procedures. Finally, we conclude the appendix with an analysis of three small question-wording experiments conducted during the survey fieldwork.

### Survey Design

The overall project design called for the development of two survey questionnaires, one for parents and the other for teachers. For reasons of comparability we incorporated as much overlap as possible. In each of the nine thematic policy areas chosen, a review of the existing international survey research literature was undertaken. While an emphasis was placed on finding recent survey designs that included both parents and teachers (e.g., Public Agenda, 2004), such designs are very rare. Twenty-seven questionnaires were identified for careful review, thirteen from Canada and fourteen from the United States (all completed within the last decade). Although other studies were identified, including surveys from the United Kingdom, Australia, and New Zealand, these twenty-seven were selected because they were recent and focused on one or more of the core policy domains. Very few of these studies systematically included both parents and teachers in designs that used parallel questionnaires. We could find no such study in Canada where there was adequate reporting of questionnaire items, survey methodology, and response rates/coverage.

From the studies selected we harvested potential questions that met several screening criteria including: relevance to policy themes, viable question wording (e.g., clarity, cross-jurisdictional comparability), and appropriateness both for parents and teachers (e.g., relatively jargon free). This generated a lengthy list of possible questions. These were winnowed down by the research team and then the core elements of possible survey questions were presented to participants in twenty different focus groups (five parent-only groups, five teacher-only groups, and ten mixed groups of parents and teachers). These groups helped us with further refinements (and the elimination of some questions). The research team then met to compose the penultimate version of the two questionnaires. This version was pretested by a market research firm, both in English and French, and both with teachers and parents, and final refinements were made.

The final survey instrument was approved by the Ethical Review Board of the University of British Columbia. The survey fieldwork was conducted between May 15 through July 19, 2005, with the bulk completed in May and June. The parent questionnaire consisted of 126 separate questions (not all questions were asked of every respondent since not all questions were applicable to every person interviewed – e.g., some questions were applicable only to parents with children in certain types of schools). The teacher

questionnaire consisted of 109 separate questions (again not all questions were asked of every respondent). Telephone interviews with parents took on average 23 minutes (std deviation: 7.0) while teacher interviews lasted an average of 20 minutes (std deviation: 6.2).<sup>1</sup> The final surveys for both parents and teachers are available upon request.

## Sampling

In order to conduct a national survey of both parents and teachers it was decided that a telephone interview methodology would be most effective. Other methods were considered but eliminated because of cost (face to face interviews), sample coverage (internet based survey), response rates (internet based surveys, mail surveys), and response integrity (mail surveys).

### i) Parent Sample

The parent sample was generated by standard random digit dialing (RDD) procedures. This involved placing slightly over 145,000 calls to randomly generated telephone numbers (with area codes and the first five digits randomly selected from working banks of telephone numbers). The sample was disproportionately stratified by region to ensure sample coverage as indicated in Table A1.

Table A1: Sample Coverage of Parents by Region

<b>Region</b>	<b>Target</b>	<b>Actual</b>
Yukon and Territories	48	49
BC	188	188
Alberta	188	188
Saskatchewan	188	190
Manitoba	188	188
Ontario	400	401
Quebec	400	404
Atlantic Provinces	400	400
National Total	2000	2008

Parents were defined as follows: A mother, father, or guardian who has one or more children currently attending a provincially recognized elementary, middle, or secondary school (or the equivalent, including public and private, as well as Protestant and Catholic) or officially enrolled in a recognized homeschooling program (where ‘recognized’ means officially registered with the appropriate provincial body that oversees homeschooling).

To ensure that we were speaking with parents of school-aged children we used the following screening methods during our first contact with a potential household. We began by asking if there were children in the household: “Are there any members of your

---

<sup>1</sup> Some interviews were split over two time periods if respondents requested that interviewers phone back. These times have not been recorded in computing these averages (45 teachers and 59 parents).

household attending elementary or secondary school?” If the response was yes, then we asked if the person on the phone was a parent or legal guardian of one of those children and if they were they were automatically selected to be interviewed. If the person on the phone was not a parent or legal guardian, then we asked to speak with the parent who had the most recent birthday. Automatically selecting the person to whom we were speaking had the effect of oversampling women who tend to answer the telephone more frequently than men. However, our pre-testing showed that the alternative of asking to speak to a randomly chosen parent or guardian significantly increased non-response rates (because the second parent was away, was busy, or was otherwise unavailable).

## ii) Teacher Sample

No national listing of teachers exists to use as a sampling frame. As a preliminary step we investigated using lists from either Teacher Federations/Unions or Provincial Ministries. For a variety of reasons we were unable to establish consistent methods for generating samples by approaching provincial bodies. This occurred for a variety of reasons including concerns over privacy and confidentiality, a requirement to vet the research prior to allowing the use of sampling lists, a requirement of teachers not to participate in any survey research, and the lack of agreement among the multiple jurisdictions that represent teachers in various regions.

After pursuing the provincial level approach for several weeks we decided to approach school boards/districts directly. The goal was to build a nationally representative data base by first randomly selecting schools (we did this by using Scott’s Directory of Canadian Schools supplemented by verification where possible from provincial ministry listings). As a second step we sought to randomly sample teachers from the listing of teachers in the selected schools.

We operationalized this procedure by first requesting permission of superintendents to approach the principals of the randomly selected schools (initial field testing revealed that approaching principals directly was unwise since they often referred the matter to the Board/District level). School selection was weighted by size (number of teachers) to ensure that we did not over-represent small schools. We then contacted school principals asking them to provide a listing of teachers who would be eligible for inclusion in the sample. For reasons of privacy we asked principals to allow teachers to opt in by providing their name and contact information to the research firm responsible for sample selection.

In the end we approached 410 Boards/Districts, representing 1,398 schools. We generated a national listing of 5,482 teachers. There is a small clustering effect introduced by potentially having more than one teacher from any single school included. This effect varies somewhat by province, and was a particular problem in Quebec where labour issues complicated our work. Importantly, even if a simple random sample of teachers were possible (and it is not), the selection of multiple teachers from the same school would occur. As we note below, we also included some teachers (N=104) who were randomly sampled as part of the RDD parent interview procedures.

The teacher sample was disproportionately stratified by region to ensure sample coverage as indicated in Table A1.

Table A1: Sample Coverage of Teachers by Region

<b>Region</b>	<b>Target</b>	<b>Actual</b>
Yukon and Territories	60	64
BC	225	230
Alberta	225	225
Saskatchewan	225	250
Manitoba	225	227
Ontario	480	486
Quebec	480	227
Atlantic Provinces	480	482
National Total	2400	2191

The serious undercoverage in Quebec (227 teachers from a target of 480) is a direct consequence of labour issues that were unfolding as the fieldwork was undertaken. We comment further on this undercoverage below.

Teachers were defined as follows: Individuals working as educators in a provincially recognized elementary, middle, or secondary school (or the equivalent, including public and private, as well as denominational schools (i.e., Protestant or Catholic)). This includes classroom teachers, librarians, learning assistance teachers, and other teachers, but does not include administrative staff (e.g., principals, vice-principals, secretaries, or other support staff). Excluded from the teaching sampling frame will be nursery aid teachers, preschool teachers, and supply teachers (which are included in Statistics Canada definition of teachers).

The vast majority of teachers were contacted at their home addresses (although some contacts were made via the school as requested by individual teachers). We began the telephone interviews by verifying that the person we were speaking with had taught during the previous 12 months in grades K-12 and that they were not currently a principal or vice-principal.

### **Sample Representativeness**

In this section we first report response rates and refusal rates for both surveys. Second, we also evaluate the quality of both samples by comparing the sample of parents and the sample of teachers to known attributes of the relevant Canadian population. One problem in making such comparisons is that the known attributes of either parents or teachers are not available in an authoritative source for the exact dates on which our samples were collected. For the parent sample, we rely on Statistics Canada data mainly from the 2001 Census as well as from the most recent General Social Survey (2003) conducted by Statistics Canada. For the teacher sample, we rely on Statistics Canada data on teachers and on data from provincial Ministries responsible for K-12 instruction. Our sources are identified for all comparisons.

## i) Parent Sample

Just over one third of all randomly dialed calls resulted in someone answering the phone. Of these 55,200 contacts, the telephone interviewers established that 14,052 households were eligible for inclusion in the sample. From the eligible households, 2,008 surveys were completed generating a response rate of 14.3% (completed interviews divided by eligible households). From the 14,052 households, 6,547 individuals refused to participate, for a refusal rate of 46.6% (refusals divided by eligible households). In many households, despite repeated efforts, we were never able to complete an interview with the selected individual.

To examine the representativeness of the parent sample we compare sample results with national benchmarks on four dimensions: gender, family income, family size, and age. These four measures are used because they represent the best benchmark data available.

*Gender:* In the parent sample, 69.4% of respondents were women. According to the most recent figures available from Statistics Canada, 54.4% of parents in households with children are female (*Census families in private households by family structure and presence of children*). This oversampling of women is in part a consequence of the decision, described earlier, to automatically include in the telephone interview, the first parent or guardian of school-aged children we were able to identify. Oversampling women is also a common outcome in market research.

*Family Income:* The average pre-tax family income of the parent sample was \$68,500 (estimated from the seven income categories used on the parent survey). In 2003, the average total income of Canadian families with children as reported by Statistics Canada was \$73,500 (*Average total income by economic family types*). The sample estimate is very close, even after adjusting the 2003 Statistics Canada figure for inflation. Respondents in the survey were asked to report “total annual family income” in one of seven categories.

*Family Size:* The following table shows the number of children for families included in the sample compared with the number of children in families who reported in the 2001 Census that they had children (*Number of Children at Home and Family Structure for Census Families*). The variation is relatively slight.

Table A3: Family Size for Families with Children (2001) and the Parent Sample (2005)

Number of Children	Families with Children %	Parent Sample %
1 Child	41.1	43.0
2 Children	42.7	39.3
3 or more Children	16.1	17.7

Age: The average age of the parent sample is 41.6 years. A perfectly comparable benchmark is not available although the 2003 General Social Survey, done by Statistics Canada, has an average age of 38.5 for parents with children between 0 and 14 living at home. The difference is not large and is in the expected direction since the sample of parents of children in K-12 should be slightly older than those parents with children between 0 and 14.

Summary: On four separate measures, two related to family characteristics and two related to personal attributes, the sample appears to meet the benchmark in three of the four cases. The fourth case, gender, is to some degree an artifact of telephone screening procedures and is adjusted for in the weighting scheme employed (see below).

## ii) Teacher Sample

From the database that was constructed we selected a sample of 4,578 eligible respondents (each of whom met the teaching definition noted above). From this group, 2,191 interviews were completed giving a response rate of 47.9% (completed interviews divided by eligible respondents). Of the 4,578 teachers we attempted to contact, 411 refused to participate in the interview generating a refusal rate of nine percent (refusals divided by eligible teachers).

The teacher sample was seriously impaired by labour issues in the province of Quebec that were occurring during the time of fieldwork. We have chosen to include the Quebec teachers in the final sample, but in all cases we have ensured that when the response pattern for Quebec differs significantly from teachers in other regions of the country, these are explicitly noted.

To examine the representativeness of the teacher sample first we compare sample results with national benchmarks on three dimensions: gender, age, and institution type (public or private). These three measures are used because they represent the best data available to use as national benchmarks. Second, we also compare two separate segments of our teacher sample, one generated via the school sampling and the other created via the random digit dialing sampling of parents.

Gender: In the teacher sample, 70.6% of respondents were women. In the most recent data publicly available (*Education in Canada 2000*), women represent 64.1% of the teachers in K-12 public and private schools. Men proved to be both harder to contact than women, and they tended to refuse participation at a slightly higher rate than women.

Age: The average age of teachers in the sample was 41.6. Statistics Canada reports the average age of public and private K-12 teachers as 42.6 (*Education in Canada, 2000*). Over the previous five year span, from 1995-96 to the most recent Statistics Canada data, there is no trend, either up or down, in the mean age of teachers.

Type of Institution: In the most recent data available (*Education in Canada, 2000*) Statistics Canada reports that, in the late 1990s, the split between the percentage of

teachers in public versus private schools was 93.4 and 6.6 respectively.<sup>2</sup> In the teacher sample the split was 87.9 versus 12.1. Examining this split more closely at the provincial level, more private schools were included in the sample for BC and Ontario, but in the other provinces the split was much closer to the provincial norm. In our analysis we adjust for this split by weighting the sample data to reflect a smaller private or independent school group.

### Comparisons Between Sampling Procedures

As we noted earlier, teachers were sampled either via a database constructed via school selections (N=2,087; hereafter School sample) or through the random digit dialing process used to generate the parent sample (N=104; hereafter RDD sample).<sup>3</sup> Comparing both the attributes of teachers and the perceptions of teachers across these two sampling procedures sheds light directly on whether or not our results are influenced by sample selection methods. One could speculate that teachers sampled via the school database might be unrepresentative because of the various stages of permission that were essential to generating this sample. In contrast, teachers included via the random digit dialing process might be distinctive for other reasons, including availability and interest. However, if there are few differences between the two samples, this suggests that response patterns are relatively uninfluenced by selection procedures. This would enhance confidence regarding the adequacy of the teacher sample.

On the measures we used in comparing teachers to national benchmarks, we find that teachers from the RDD sample are of similar age (42.4 vs. 41.6) and gender (28.8% vs 29.4%) relative to the School sample. However, we anticipated that the RDD sample would have fewer teachers who taught at private or independent schools given the suspected overcoverage from these types of schools in the School sample. Exactly the opposite occurred. In the RDD sample just over 20% of the teachers were from the private and independent sector, as compared to closer to 10% in the School sample from that sector. The numbers are small; ten more teachers than we expected in the RDD sample teach in the private and independent sector.

Teachers in the RDD sample, as compared to the School sample, are no different with respect to fulltime – parttime status, the size of the schools at which they teach, or the number of years of teaching experience they have. A slightly greater proportion of the teachers in the RDD sample teach at the secondary school level as compared to the School sample.

---

<sup>2</sup> In our parent sample, 8.1% of parents report having one or more of their children in a private or independent school. If the private/independent sector has grown in the last five years, as we suspect it has, then this is a realistic estimate of the split between public and private/independent.

<sup>3</sup> Teachers selected in this method did not have to be parents. In the parent survey the very first question asked of whomever answered the phone was “Does anyone in the household teach in the K-12 system?” If the response was yes, the telephone interviewer asked to speak with that person and immediately switched to the teacher questionnaire.

When we contrast the RDD and School samples on teacher perceptions, again we find few differences. First, when asked to give schools a grade from A through D, or Fail, the mean grades are virtually identical. Second, on an array of other attitudinal measures, from views on standardized tests through the use of annual report cards for schools, there are only minor differences between teachers selected via or RDD or Schools.

In summary, the comparisons to available national benchmarks and the internal comparison of two different sampling methods both suggest that the sample is representative of Canadian teachers. The sample is weakest for Quebec. The unweighted sample also contains a disproportionate number of women and teachers from private and independent schools.

### **Sample Confidence Intervals**

With both the parent and teacher samples exceeding 2,000 respondents we can be confident that 19 times out of 20 the sampling errors are  $\pm 2.2$ . This means that in 95% of all cases our estimates of percentages are likely to vary by only plus or minus 2.2 percent for each of the full samples. When we combine the two samples, as we do when we compare results for teachers and parents, the confidence intervals are even smaller since the sample size exceeds 4,000 in these cases.

We do present a few results at a regional level but in all cases where this happens we remind readers of the smaller sample sizes on which these estimates are based.

### **Sample Weights**

The findings reported in the main body of the report are based on weighted data. Weighting is used for two purposes. First, since the sample was weighted disproportionately by province to ensure sufficient minimal coverage in the major regions of the country, weights are employed to scale the results to the national level where each respondent from each province contributes their proportional share to the Canada-wide results. Second, since each of the samples either under or over represents certain groups, the national sample is weighted to redress this balance as appropriate.

#### **Parent Sample**

The parent sample over-represents women and therefore weights have been used to ensure the national level results reflect the appropriate balance of women and men who are parents or guardians of school-aged children. As well, since we intentionally over-sampled parents in some of the smaller regions (e.g., Atlantic provinces) and under-sampled them in others (e.g., Ontario), the national parent sample is therefore also weighted by region.

## Teacher Sample

The major weighting factor in the teacher sample is an adjustment for the deliberate over- and under-sampling of different regions of the country (similar to what was done in the parent sample). In addition, minor adjustments in the teacher sample have been introduced to rebalance the gender ratio and the type of institution at which teachers work. This effectively means a slight downweight for female teachers and a slight downweight for teachers employed in independent/private schools. These minor weighting adjustments ensure the teacher sample better reflects the national distribution of teachers by gender and institutional type.

We have examined a variety of different relationships between key variables (e.g., age and grades given to schools), and the results are consistent whether weighted or unweighted data are used. Nevertheless, the results are presented using weighted data since this more accurately reflects the distributions of key variables at the national level.

## Question Wording Experiments

The amount of survey research being conducted in education is increasing (indeed one oft-repeated reason people gave for refusing to participate in this study was their sense of being “over-surveyed”). Especially as various educational bodies attempt to provide enhanced measures of accountability, surveys of the public, of parents, and of teachers proliferate. The value of the evidence derived from these surveys is sometimes contested, especially on two grounds. A first criticism concerns response rates and the potential for low rates to produce biased results. We have directly addressed this worry above. We are confident that our sample meets accepted standards of being nationally representative both for parents and teachers.

A second concern has to do with the accuracy of survey responses. Here the challenge concerns the questions asked, and especially the worry that questions can be misleading or biased. Question wording is an inexact science and while general rules of thumb guide word selection, the final wording typically emerges from a consensus of those responsible for drafting the questionnaire. Given that there is no exacting logic about what precise words to use we examined how robust questions might when different word choices were made. We also wanted to use the opportunity provided by the parent and teacher samples to see if for one or the other group, word choice might be more or less robust.

We ran three experiments within both the parent and teacher surveys, choosing to ask one randomly chosen segment of the sample one question, and the other segment a different variant of the same question. In all cases the questions differed in only one word. Based on our focus groups, we chose word pairs that had sparked discussion. The three wordings were as follows, with the word changes underlined:

A1: Government funding should follow the child, regardless of where he or she is enrolled ...

A2: Government tax money should follow the child, regardless of where he or she is enrolled ...

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree
- Don't Know/Refused

B1: Is the ability of teacher unions to influence what happens in schools ...

B2: Is the power of teacher unions to influence what happens in schools ...

- About right
- Too strong
- Not strong enough
- Don't Know/Refused

C1: Provinces should administer standardized assessments at key stages of schooling ...

C2: Provinces should administer standardized tests at key stages of schooling ...

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree
- Don't Know/Refused

In each case we felt that the second variant was on the edge of being overly provocative even though the basic meaning of each question is very similar, if not identical.

Does wording matter? Yes – at least certain results depend upon wording. In five of the six experiments the response patterns for the individual items differ significantly by question wording. Table A-4 (next page) summarizes the findings, showing relevant percentages for both parents and teachers.

As the first panel shows, both parents and teachers are more likely to strongly agree that government funding should follow students, but are somewhat less likely to support government taxes doing likewise. The differences are not large – about five percent for parents and eight percent for teachers – but the type of wording used influences the distribution of the results (and for those interested, both differences are statistically significant at the .05 level using Chi Square). The second panel on unions suggests that for a majority of teachers, union ability to influence what happens in schools is ‘not strong enough’ although about ten percent fewer teachers believe union power to influence what happens in schools is ‘not strong enough.’ Only in the case of parents and the question about union ability or power, are the results intransient to wording effects

(the three percent difference here being relatively small in the context of the other response categories on this question). A first conclusion, then, is that wording does matter.

Table A-4: Results of Wording Experiments by Question, Parents and Teachers  
(in Percent)

Response	Government	
	Taxes	Funding
Parents - Strongly Agree	58.3	63.6
Teachers - Strongly Agree	53.8	61.9
	Unions	
	Ability	Power
Parents - Not Strong Enough	29.0	26.0
Teachers - Not Strong Enough	55.6	46.8
	Standardized	
	Tests	Assessments
Parents Strongly Agree	44.2	49.5
Teachers Strongly Agree	23.5	21.1

Note: Parent N = minimum 1651; Teacher N = Minimum 1971; weighted data.

Notice a second pattern, however. The relationships between variables do not depend upon question wording. It does not matter how you ask parents or teachers about tests or assessments, or about union ability or power, the two groups differ. For example, almost half of all parents support some form of standardized testing or assessment, whereas fewer than a quarter of teachers support such measures.

A second example is illustrative here. Does the relationship between a teacher's years of experience (as an example) and their views on union influence depend upon whether the question about unions is phrased as ability or power? No. The basic bivariate pattern repeats for each version of the question suggesting that the relationship is robust enough to be uninfluenced by exact word choice.

Two significant lessons emerge from this simple experiment. First, interpreting the results of any single questionnaire item in isolation from other information is problematic because without benchmarks or contextual detail, results are transient to a range of factors, including the exact wording of the question. Second, data patterns and relationships between different variables are often less influenced by question wording. Indeed knowing that a relationship between two variables holds even though the question is asked slightly differently to different groups, enhances the robustness of the findings.