ANALYSIS OF POTENTIAL NONRESPONSE BIAS

J. Michael Brick, Westat and Jonaki Bose, National Center for Education Statistics
J. Michael Brick, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Response Rates, Nonresponse Adjustments, Simulating Nonresponse, Panel Response

1. Introduction

This paper summarizes the major findings of a nonresponse bias analysis conducted on the base year data collection of the Early Childhood Longitudinal Study: Kindergarten Class of 1998-99 (ECLS-K). The analysis examines the potential for bias in the estimates due to nonresponse and looks at the effect of the nonresponse weighting adjustments that were made to reduce the bias.

The ECLS-K is a study conducted by the National Center for Education Statistics in the U.S. Department of Education. Westat collected the data for the survey. The survey focuses on children's early experiences, beginning with kindergarten and proceeding through fifth grade. It is a nationally representative sample with 21,260 children participating in the base year (National Center for Education Statistics, 2001). Overall, six waves of data collection are currently planned: two in kindergarten (fall and spring), two in first grade (fall and spring), third grade (spring) and fifth grade (spring). Burke, Lê, and Brick (1998) describe the sample design for the ECLS_K. This paper covers data collected in the base year, the fall and spring of kindergarten (1998-1999).

In the base year, data were collected from a variety of sources: children, parents, teachers, and schools. Children were administered direct student assessments in reading, math and general knowledge. Parents participated in a telephone interview that included topics such as family structure, parental involvement, and child care history. Teachers provided data about their classrooms, instructional practices, beliefs, and background, as well as information on the social and academic performances of sampled children. These collections were done in both the fall and the spring. In addition, in the spring of kindergarten, school administrators were surveyed on subjects such as the characteristics of the school, students enrolled in the school, school facilities, teacher and staff characteristics, school policies, principal characteristics and school climate.

Each level and source of data required for the ECLS-K resulted in an opportunity for nonresponse. For example, not all school districts and schools agreed to participate and no data could be collected in this case. Within schools that did participate, the response rates for children, teachers, parents, and school administrators differ for a variety of reasons.

Because each source had different reasons for responding or not responding, methods to reduce nonresponse were tailored to the source. Respondents were provided a toll-free number they could call to verify the legitimacy of the survey. Teachers, schools and parents were sent newsletters describing study results between rounds of data collection.

Each child was provided an ECLS-K multicolored pencil upon completion of the child assessment. Children participating in the study are sent birthday cards. In the parent interview, response rates were increased by repeated phone calls and extensive follow-up. For households without a telephone, face-to-face interviews were conducted to ensure inclusion in the survey. Upon completion of the interview, parents were mailed a thank-you letter and a copy of a Department of Education publication "Learning Partners – A Guide to Educational Activities for Families."

Teachers were given $5 for each child-level questionnaire they completed in the fall of kindergarten. This was increased to $7 in the spring of kindergarten. In the case of the self-administered teacher and school questionnaires, the response rates at the planned end of data collection (corresponding to the time schools closed for the summer) were not as high as hoped. To increase the response rates, the data collection period was extended for a few additional months after school started back up in the fall.

Despite these and other methods to encourage cooperation, nonresponse did occur. Since we are dealing with unit nonresponse, adjustments in the sampling weights were made for each type of respondent to compensate for the nonresponse. For example, the sampling weights for teachers were adjusted for nonresponse at the school and teacher level. These weighting adjustments were made after all the attempts to gain the sampled units cooperation were exhausted. The goal of the nonresponse adjustments is to reduce the bias in the estimates due to nonresponse and to make analysis of the survey responses relatively simple. (National Center for Education Statistics, 2001) gives a full discussion of the nonresponse weighting adjustments.

Nonresponse in a survey with so many sources of data and relationships between the sources is a very complex phenomenon. Evaluation of the effect of nonresponse is even more difficult because only limited data are available for the nonrespondents. To deal with these problems, multiple methods are used to investigate potential nonresponse bias. The sections that follow describe different methods of examining aspects of nonresponse and the potential effect on the estimates. The final section synthesizes the findings and speculates on the overall nature of nonresponse bias in the ECLS-K estimates from 1998-1999. The more extensive nonresponse evaluation report (National Center for Education Statistics, 2002) summarized in this paper contains many tables that could not be presented here due to space limitations.

2. Examination of Response Rates

The first evaluation approach is the examination of response rates. While the level of nonresponse does not necessarily translate to bias, large differences in the response rates of subgroups serve as indicators that potential biases may exist. The often-used expression for the bias of the mean from a simple random sample is valuable for expressing this relationship. The bias is

$$b(\bar{y}_r) = (1-r)(\bar{y}_r - \bar{y}_{nr});$$

where the subscripts *r* and *nr* denote respondents and nonrespondents, respectively, and (1-*r*) is the nonresponse rate. So, if the response rates for high- and low-income children were very different, any difference between the means of the respondents and nonrespondents would result in a large bias.

While this approach may indicate the potential for bias, it is limited because it does not deal with any nonresponse adjustments made to reduce the bias. Methods presented later go further in this direction. Another limitation of this approach is that response rates can be calculated only for those subgroups where the subgroup characteristics are known for both the respondents and nonrespondents. In the ECLS-K this information is taken either from the data on the sampling frame for school level characteristics or from data collected at the schools for child and teacher level characteristics.

Prior to discussing nonresponse in the base year of the ECLS-K, it is useful to establish some terminology. Completion rates refer to the percentage of participating units at each stage of sampling and are calculated for different ECLS-K components and questionnaires. Response rates refer to the overall percentage of participation in the study and take all stages of sampling into account. In the ECLS-K, the response rate is a product of the school response rate and the completion rate of a given component. Completion rates help identify differences within subgroups at the same level, while response rates describe the broader picture but confound the sources of bias. All completion and response rates presented are computed using weights adjusted for the probability of selection.

The full report contains response rates for almost all the characteristics available from the sampling frame and they are broken down by characteristics of the school. Here we concentrate on the more exceptional rates that might indicate potential bias. The lowest completion rates for the ECLS-K in 1998-99 are at the school level. In the fall, 69 percent of the sampled schools participated in the study. The school completion rates were higher in the spring of kindergarten when more schools agreed to participate, bringing the school completion rate to 74 percent.

The school participation rates varied depending on type of school. Some of the largest differences in rates are: 78 percent of schools in large cities participated compared to 68 percent in urban fringes of these large cities; the Catholic school completion rate was 83 percent as compared to 56 percent for non-Catholic private schools.

Completion rates at the child level were calculated for all six data collection instruments or components: child assessments, parent interviews, teacher questionnaires A, B and C, and school administrator questionnaires (SAQ). No major patterns or differences in completion rates are apparent among different subgroups, other than by school type (Table 1).

For the SAQ, completion rates at a school level are relatively consistent with only one exception. Public schools with high percent minority enrollment had lower completion rates than other types of schools. This difference also shows up at the child-level (Table 2).

One of the benefits of this type of analysis for a longitudinal survey is that it can be used to improve future data collection in the survey. Based on the differences by percent minority, special efforts were made for high minority schools in the first grade data collection to improve their SAQ completion rates. While the difference in the completion rates between low and high minority schools could not be completely eliminated, the efforts reduced the difference by over 20 percentage points in the first grade collection.

Table 1. Spring 1999 ECLS-K kindergarten child-level completion rates, by component and type of school

Component	Public school children	Private school children
Child Assessment	88%	90%
Parent Interview	83%	88%
Teacher Questionnaire B	90%	95%
School Administrator Questionnaire	85%	90%

Table 2. Spring ECLS-K kindergarten child-level school administrator questionnaire completion rates, by percent of minority students and type of school

Percent minority	Public school children	Private school children
0-10 percent	92%	98%
11-49 percent	93%	95%
50-89 percent	79%	82%
90-100 percent	61%	80%
Unknown	100%	78%

Overall, the analysis shows that type of school (public, Catholic, non-Catholic religious, other private) is the major source of variation in the completion and response rates. As mentioned earlier, this approach does not take any weighting adjustments into account. In the ECLS-K, the weight adjustments specifically address variation based on type of school. Consequently, the effect of the adjustments should be to reduce nonresponse bias, especially for characteristics correlated with type of school.

3. Comparison of Sample and Frame Estimates

The second approach for examining the potential for nonresponse bias in the estimates from the ECLS-K involves comparing sample estimates from the responding schools and children to the population values computed from the sampling frame. Clearly, only variables on the sampling frame can be used in such comparisons. The strength of this approach is that any differences are due solely to sampling and nonresponse error. The weights used in this comparison are based on the probability of selection, with no nonresponse adjustments.

The difference between the sample estimate and population value from the frame was calculated and a 95 percent confidence interval was estimated for the difference for school- and child-level measures. If the 95 percent confidence interval contained zero we assumed the difference between the sample estimate and population value was not statistically significant.

The Common Core of Data for public schools and the Private School Survey for the private schools are the two main frames used in the ECLS-K. The few schools selected from other sources were not included in this analysis.

The school characteristics examined were: type of school, school affiliation, type of locale, region, and kindergarten type. Overall, in each round 27 differences were computed. The only statistically significant difference is in the percentage of non-Catholic religious schools (13 percent in the frame and 10 percent in the sample).

At the child level the characteristics examined included: type of school, type of locale, region and race/ethnicity distribution of children in all grades. Again, only one difference was statistically significantly different from zero, the percentage of children in non-Catholic religious schools for both the fall and spring kindergarten sample.

Thus, in over 94 comparisons, only 4 of the differences were statistically different from zero. This is close to the expected number if the comparisons were sampled from a distribution with a mean of zero. Thus, the analysis does not show a systematic bias due to nonresponse. If nonresponse adjusted weights had been used, it is likely that the differences between the frame and sample estimates would have been further reduced.

4. Comparison With External Data Sources

The third approach used to investigate the potential nonresponse bias is to compare estimates from the ECLS-K to estimates from other surveys with similar items. Large differences may indicate potential bias and the need for further study. However, differences cannot be solely attributed to nonresponse bias, since there are many other possible sources of the differences. For example, estimates from different surveys may not be comparable because of coverage disparities, time periods that are not the same, differences in question wording, context effects, and a host of other nonsampling error sources. Despite these severe limitations, differences in estimates serve to alert users to potential concerns and may facilitate uncovering important issues. The estimates from the ECLS-K and the other surveys were computed with the fully adjusted weights used for the specific surveys.

Since the base year of the ECLS-K has nationally representative samples of kindergarten children, kindergarten teachers, and schools containing a kindergarten in 1998-99, it is possible to compare estimates at all three levels to estimates from other surveys. The surveys that provided estimates that could be compared to the ECLS-K estimates, at least one of these levels, are the 1993-94 Schools and Staffing Survey (SASS), the Fast Response School Survey (FRSS), the National Household Education Survey (NHES), and the Current Population Survey (CPS).

The SASS and ECLS-K estimates were restricted to public and private schools with a kindergarten, but not having a high grade of kindergarten, to create comparable estimates. The question wording of most of the estimates compared were also very similar. The estimates of percent of schools by school type, and average school enrollment and attendance are very similar. Differences are found in the estimates of the percent of Catholic schools with Title I programs and the percent of private schools with gifted and talented programs. Some of these differences may be due to the time periods for the two surveys. The SASS data were collected in 1994-95, four years before the ECLS-K.

The FRSS teacher survey of 1993 was used as a source of teacher-level estimates of public school teacher characteristics and opinions for comparison purposes. The estimates from the FRSS and the ECLS-K show both similarities and differences. The ECLS-K estimates more teachers in urban fringes than the FRSS. There are also differences in the number of teachers in full and half day programs and in the class size of kindergarten teachers. The differences are consistent with current trends that have more teachers in full day programs and teachers teaching smaller classes, but it is not clear that the differences are merely due to the passage of time and changes in the kindergarten educational policies. The ECLS-K and FRSS estimates are almost identical in teacher characteristics by race, mean number of years teaching kindergarten and major/certification in early childhood education. The percent of minority pupils in elementary schools are different between the two studies. The higher ECLS-K numbers may be explained by the actual increases in minority student populations. There are also some differences in the opinion items, but no major inconsistencies.

The NHES is a source for comparing estimates based on parent's responses. The NHES was conducted several times in the 1990s, but most of the comparisons presented here are based on estimates from the 1995 and 1999 surveys. There is a relatively large difference (6-percentage points) between the 1995 NHES estimate of the percent of children who attend full-day kindergarten programs and the ECLS-K estimate. In addition, some large differences by family type are also observed. The estimates also differ for the following characteristics: the distribution of children by the time they entered kindergarten, the number of times family members read to the child, the child's participation in non-parental childcare, parent's participation in their child's school, how well the parents think their child's school does various tasks, and the parent's participation in different activities with their child. Some of the differences may be due to how the child's home primary home language was classified, the way the questions were worded and the response options provided to the respondents. In cases such as participation in school activities, even though the ECLS-K percentages are lower than the NHES-99, the differences are constant across different subgroups defined by child's race/ethnicity, mother's level of education, mother's employment status, family type and primary language spoken at child's home. On the other hand, estimates such as percentage of kindergarten children in nonparental care arrangements differ in the overall percentage and by subgroup.

The CPS and NHES-99 were used to compare overall household level estimates. The October 1998 CPS estimates of the percent of children by race/ethnicity are compared to the ECLS-K distribution. The ECLS-K has higher percents of Hispanics and fewer black children who were first-time kindergartners. There are several possible reasons for these differences. The ECLS-K has a 'more than one race' category that the CPS does not, so there are questionnaire differences. In addition, the CPS is calibrated to the 1990 decennial census, which may result in lower sensitivity to changing minority percents. The recent releases of data from the 2000 census by race and ethnicity support this conjecture. The NHES-99 estimates by household income categories are

different from the ECLS-K estimates at the lower and upper ends of the distribution. The ECLS-K estimates more households at the lower income categories and the NHES-99 estimates more households at the higher income categories. This difference may be due to an over-representation of high-income households in the RDD NHES survey, rather than nonresponse bias in the ECLS-K survey.

While we concentrated on the differences in this summary, it is clear that not all of the estimates from the ECLS-K and the other surveys are comparable. The problem with this type of comparison is that the reasons for the differences may not be related to nonresponse bias. Such comparisons are very limited for investigating nonresponse because all of the sources of each survey are confounded in the differences and there is no way to attribute differences to single sources from one of the surveys.

5. Comparison of Adjusted and Unadjusted Estimates

The fourth approach to evaluating bias in the base year ECLS-K data is to compare ECLS-K estimates that include the adjustments for nonresponse to estimates made from weights that do not have any nonresponse adjustments (base weights). The main goal of the approach is to examine the effect of the nonresponse adjustments on the estimates.

Two nonresponse-adjusted weights are used in this effort. The first is the child-parent-teacher (CPT) weight. This weight considers all those children who are missing any of the three possible instruments (a child assessment, a parent interview, or a teacher questionnaire) as nonrespondents and adjusts accordingly. This weight is the most restrictive of the ECLS-K cross-sectional weights. The second is the parent weight that considers children with a parent interview as respondents, irrespective of the other instruments. Since the parent response rate is the lowest of the three instruments, this is the second most restrictive weight. The unadjusted estimates use the base weights and only data from original schools. As a result, data from substitute schools are treated as part of the nonresponse adjustment.

Large differences between the unadjusted and adjusted estimates may indicate the potential for nonresponse bias, but this approach has the same limitation as the response rate approach in that it does not account for the weighting adjustments. If the differences between subgroups are associated with characteristics that are used in the nonresponse adjustment process, then this approach will not reflect this fact.

Child assessment scores, social rating skill scores from the parents and the teachers, and the distribution of children by family characteristics are compared for two rounds of data collection using the different weights. In most cases the differences are less than one percent no matter which of the adjusted weights is used. The largest difference is only two percent. Thus, no important differences for the characteristics studied were found for either of the adjusted weights.

6. Simulating Nonresponse

As noted earlier, in the base year of the ECLS-K data for each child were collected from different components: the child assessment, the parent interview, the teacher questionnaire, and the school administrator questionnaire. Cross-sectional and longitudinal weights were developed depending on which of the three components were completed for the children. The variability in the number of completed components for the children provides a realistic basis for simulating nonresponse without having to make contrived assumptions. This is the rationale for the final evaluation technique.

If the child assessment in a round was completed, then the child had a positive C-weight for that round. Similarly, if a parent interview for a sampled child was completed, then a P-weight was assigned for that round. As described in the previous section, a CPT-weight was assigned to each child in a round if all three components (the child assessment, parent interview and teacher questionnaire) were completed. The final weight we consider is a panel CPT-weight or CPT-p weight. This weight is assigned only for children with all three sources of data for both round 1 and round 2.

Having multiple definitions of response provides the opportunity to estimate the same characteristic using different sets of respondents and the corresponding nonresponse adjusted weights. For example, in round 1, a child assessment-based estimate could be computed using the C1, the CPT-1 and the CPT-p weights. The differences between the estimates are only due to the differential nonresponse. Thus, the strength of the simulation method is that it provides a very direct estimate of the bias due to differential nonresponse in the ECLS-K.

For round 1, the difference between an estimate from the child assessment using the C1 weight (based on all the children with these data) and the CPT-p weight (the smallest subset of responding children with these data) measures the differential nonresponse bias associated with the observed response pattern. The CPT-1 weight is also used to estimate these child assessment characteristics and the nonresponse bias

associated with this set of respondents. For parent interview estimates, the weights P, CPT and CPT-p are used. For teacher questionnaire data, the CPT and CPT-p weights are used. The estimates are computed separately for both round 1 (using C1, P1, CPT-1 and CPT-p weights) and round 2 (using the C2, P2, CPT-2 and CPT-p weights).

Three methods of evaluating the differences in the estimates are considered. First, we just examine the size of the differences. Second, we compare the size of the difference to the standard deviation of the estimate, a metric used in substantive analysis of these estimates. Third, we compare the size of the difference to the standard error of the estimate to assess the bias in relationship to the sampling error.

In round 1, the largest difference in the absolute size of the estimates in terms of percentages is 1.5 percent in the child assessments, while in round 2 it is less than 2.5 percent (Table 3 has some round 2 estimates). In terms of standard deviations none of the differences were more than 0.05 standard deviations. Because of the large sampling sizes, the standard errors in the ECLS-K are very small, and the differences relative to the sampling errors are greater than one. When the ratio is greater than one, the differential nonresponse bias contributes more to the mean squared error (MSE) than the variance. Ratios of greater than one are common for the overall estimates, but for subgroups the ratios are less than one (the sample sizes and consequently the sampling errors are larger for the subgroups and the size of the bias is relatively constant).

The differences in the estimates from the different weights and respondents based on data from the parent interviews and the teacher questionnaires are relatively small (Table 4 has some estimates from the teacher questionnaire data). These findings suggest the potential for differential nonresponse bias is low after the adjustments are made to the weights.

Even though the differences are small, a pattern may be present. Children with more completed components have higher mean scores in all of the assessments for both the fall and spring of kindergarten. Table 3 illustrates this for the spring kindergarten direct assessment scores. A similar pattern also is suggested in the teacher social rating scales. Children with all six sources of data (CPT-p respondents) tend to have higher pro-social score and lower anti-social scores than children for whom only the three components for round 2 (the CPT-2 weights) are required (Table 4). The patterns in the child assessment and teacher questionnaire data are suggestive, but not definitive. No patterns are evident for estimates from the parent social rating scales. The results of these patterns of missing data can be used in nonresponse adjustment weights for future rounds.

7. Summary

The analysis used five different approaches to examine the potential for nonresponse bias from the ECLS-K data collected in the base year. No method gave a strong indication that the ECLS-K estimates are subject to substantial nonresponse bias. Some areas of potential bias were identified such as schools with a high percentage of minority children in the school administrator questionnaire. Another potential source of nonresponse bias identified was the situation in which children with fewer completed components tended to have lower assessment scores and less positive social ratings from their teachers.

Methodologically, the nonresponse bias analysis took advantage of both the longitudinal nature of the survey and the multiple components of the survey. The longitudinal feature provides an opportunity to address potential bias in subsequent rounds, as noted in the high minority data items for the school administrator questionnaire. Similarly, the persistence of lower assessments scores and less positive social ratings among those children with more components completed will be monitored in the next round of data collection and these data may be used in weighting adjustments. The presence of multiple components was exploited to examine the effect of the nonresponse and nonresponse adjustments using actual patterns of missingness. This provides a much more realistic and assumption free analysis device rather than relying on assumptions such as missing at random.

8. References

Burke, J., Lê, T., and Brick, J.M. (1998). Sample design issues for the base year of a longitudinal survey of kindergarten children. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 715-720.

National Center for Education Statistics. (2001). *ECLS-K Base Year Public-Use User's Manual*. (NCES 2001-029).

National Center for Education Statistics. (2002). *ECLS-K Base Year and First Grade Methodology Index* (unpublished compilation).

Table 3. Spring ECLS-K kindergarten mean assessment scores, by type of response and weight

Mean assessment score	C2 weight	CPT-2 weight	CPT-p weight
Reading scale score	31.6	32.0	32.1
Mathematics scale score	27.4	27.7	27.8
General knowledge scale score	26.8	27.1	27.3

Table 4. Fall kindergarten mean teacher social rating skill scores, by type of response and weight

Mean social rating skill score	CPT-1 weight	CPT-p weight	Difference
Approaches to learning	2.97	2.99	-0.02
Self-control	3.08	3.09	-0.01
Interpersonal	2.97	2.98	-0.01
			0.01
			0.01