EXAMINING ADDITIONAL NONRESPONSE BIAS INTRODUCED THROUGH ATTRITION

Jonaki Bose and Jerry West, National Center for Education Statistics
Jonaki Bose, National Center for Education Statistics, 1990 K St. NW, Washington DC 20006

Key Words: Response Rates, Nonresponse Bias, Simulating Nonresponse, Panel Attrition

1. Introduction

This paper[1] summarizes the major findings from a nonresponse bias analysis of the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) first grade data. The analysis examines the potential for bias in estimates from the ECLS-K due to attrition (nonresponse) between the base year (rounds 1 and 2) and first grade (rounds 3 and 4). This work builds on a base year nonresponse bias analysis (Brick and Bose, 2001).

The ECLS-K is conducted by the U.S. Department of Education, National Center for Education Statistics. Westat collected the data for the study. The study focuses on children's early school experiences, beginning with kindergarten and continuing 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, there are six waves of data collection: two in kindergarten (fall 1998 and spring 1999), two in first grade (fall 1999 and spring 2000), and one at third (spring 2002) and fifth (spring 2004) grades. Burke, Lê, and Brick (1998) describe the sample design for the ECLS-K.

In fall and spring first grade, data were collected from a variety of sources: children, parents, teachers, and schools. Children were assessed in reading, mathematics and general knowledge. Parents participated in a telephone interview that included topics such as family structure, parental involvement, and child care. In addition, in just the spring of first grade, school administrators were surveyed on subjects such as characteristics of their student body, facilities, teachers and administrative staff, policies, and climate. Also, in spring first grade, teachers answered questions about their classrooms, instructional practices, beliefs, and background, as well as provided information on the social and academic performances of sampled children.

Despite targeted attempts to improve response rates across the different components, nonresponse did occur. Given the presence of multiple components, each child might have one or more missing components. Adjustments in the cross-sectional and longitudinal sampling weights were made to compensate for the nonresponse in the different components and waves. 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.[2]

Nonresponse in a survey collected over multiple waves, with so many sources of data is a very complex phenomenon. Evaluation of the effect of nonresponse is even more difficult because only limited data are available for the nonrespondents. This problem is especially acute in the base year, where the only information on nonrespondents is information from the frame. However, there is much more detailed and complex information available for first grade nonrespondents. This is because only base year respondents were eligible for the first grade data collections. Thus, this analysis differs from the base year analysis in that it examines additional attrition bias between the base year and first grade using additional survey-based data from the base year.

The sections that follow describe the different methods used to examine aspects of attrition and its potential effect on the estimates. However, prior to discussing the first grade nonresponse bias analysis, the next two sections briefly describe the results of the base year nonresponse bias analysis, and the characteristics of the first grade sample.

2. Summary of the Base Year Analysis

The analysis of nonresponse bias in the base year data included five different approaches: 1) examination of response rates, 2) comparison of estimates based on the ECLS-K sample to population estimates using characteristics from the frame, 3) comparison of ECLS-K survey estimates to estimates from other surveys with similar items, 4) comparison of ECLS-K estimates using nonresponse adjusted weights to estimates using unadjusted weights (base weights), and 5) nonresponse simulations using weights based on increasing levels of nonresponse.

The nonresponse bias analysis was designed considering both the longitudinal nature of and the presence of multiple components in the study. The longitudinal feature provided an opportunity to address potential bias in subsequent rounds. The presence of multiple components was utilized to examine the effect of the nonresponse and nonresponse adjustments using actual patterns of missingness. This analysis aims to utilize some of the same features. In the base year analysis, no method gave a strong indication that the ECLS-K

[1] This paper is intended to promote the exchange of ideas among researchers and policy makers. The views expressed in it are part of ongoing research and analysis and do not necessarily reflect the position of the U.S. Department of Education.

[2] A full discussion of the nonresponse weighting adjustments in first grade is found in an unpublished report (National Center for Education Statistics, 2002a).

estimates are subject to substantial nonresponse bias. One area of potential bias was identified where children in schools with a high percentage of minority students were less likely to have a completed school administrator questionnaire.

3. Characteristics of the First Grade Samples

Only children who were respondents in the base year were eligible to participate in the fall and spring first grade. Due to the high proportion of movers between base year and first grade, not all movers were followed. Movers were subsampled, and their weights adjusted to represent all movers. In addition, the fall first grade data collection was a subsample, including approximately 30 percent of the base year respondents. The spring first grade data collection included all eligible base year respondents. Also, in the spring of first grade, children were included in the data collection who were not in kindergarten in 1998-99, but were in first grade in 1999-2000 (e.g., first grade children entering the country or repeating first grade in 1999-2000), so that the spring first grade sample was representative of all first graders (National Center for Education Statistics, 2002b). First grade children introduced into the sample in this way are not included in this analysis.

Three main approaches were used to examine potential nonresponse bias associated with sample attrition between the base year and first grade. We examined response rates for the different instruments, compared respondents to nonrespondents to identify differences in their characteristics, and examined estimates based on different weights created to adjust for different sources of nonresponse.

4. Response Rates

As in the base year, response rates for the different components (e.g., teacher, parent, child assessment) were examined for variations from the base year and to identify new sources of nonresponse. 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 term for the bias of a mean is valuable for representing 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 were low, any difference between the means (e.g., mean household income) of the respondents and nonrespondents would result in a large bias when estimating that mean. While this section examines the extent of nonresponse, the following section examines the differences between respondents and nonrespondents, and the corresponding effect on the bias.

For the first grade we focused primarily on completion rates. In this case, completion rates refer to the percentage of children having completed data for each of the different instruments without considering nonresponse at earlier stages of sampling. Child-level weighted completion rates were calculated for each data collection instrument: child assessment, parent interview, teacher questionnaire, and school administrator questionnaire (SAQ).

Despite overall decreases in completion rates, there were no major differences from the base year in the completion patterns for characteristics such as gender, year of birth, school size, urbanicity, school type, and region. In the base year, schools with a high percent of minority students had lower completion rates for the school administrator questionnaires. In the first grade, efforts were made to increase completion rates in these high minority schools. High minority schools were identified, and specifically targeted by the interviewers. If the school administrator was reluctant to fill out the questionnaire, interviewers were asked to administer a set of analytically critical items, in order to at least have some data on these schools. This targeted effort resulted in higher completion rates for SAQs in the high minority schools, even though the completion rates dropped overall (table 1).

One of the main reasons for the overall drop in completion rates was the high percent of movers. Generally, most movers went to schools that were not part of the original ECLS-K sample, and these schools had less of a vested interest in participating. Thus, movers contributed to nonresponse both because they were more likely to be unlocatable, and the schools they attended were less likely to participate. And, even if the school participated by allowing the child to be assessed, the teachers and school administrators were less likely to complete their questionnaires. Table 2 shows some of the response rates associated with movers and their effects on the overall rates. While this approach may suggest the potential for bias, it is limited because it does not include any nonresponse adjustments that were made to reduce bias.

5. Comparison of respondents and nonrespondents

As mentioned earlier, the two components of nonresponse bias are the level of nonresponse associated with an item, and the magnitude of the difference between respondents and nonrespondents for that item. The second approach examined differences between respondents and nonrespondents to the entire parent interview and direct child assessment.

In the base year, the only data we had on nonrespondents was what was available from the frame. We were restricted to very basic characteristics such as race/ethnicity, gender, school type, and school enrollment. In the first grade, however, in order to be

eligible to participate, a child had to have been a respondent in the base year.[3] Thus, we have data other than frame data, for all eligible first grade children regardless of whether they were respondents in first grade. As a result, in this study, we were able to use the data collected in base year to examine characteristics of first grade nonrespondents and respondents. Using these data we were able to answer three main questions:

- Are there differences between respondents and nonrespondents?
- If yes, do these differences contribute to potential bias?
- If yes, are these estimates of potential bias substantively significant?

The first step was to examine whether there were differences between respondents and nonrespondents on base year characteristics such as race/ethnicity, gender, urbanicity, poverty, socio-economic status, parental education, parental occupation, household food security, and school type. In addition, respondents and nonrespondents were examined in terms of their total number of completed survey instruments in the base year, and their base year math, reading and general knowledge assessment scores. We compared respondents and nonrespondents for four different groups. The groups were defined based on whether a child had completed data from the:

- Fall First Grade (round 3) Child Assessment
- Fall First Grade (round 3) Parent Interview
- Spring First Grade (round 4) Child Assessment
- Spring First Grade (round 4) Parent Interview

For example, for rounds 3 and 4, children who completed a direct child assessment were compared to children who were eligible, but did not complete a direct child assessment. Similarly, for both rounds 3 and 4, we compared children with parent interview data (respondents) to children with no parent interview data (nonrespondents). These subgroups are important, as final nonresponse adjusted weights are created for each of the four groups.

Separate fall and spring first grade base weights, reflecting the fall and spring first grade selection probabilities, and their replicates were used in the estimation of characteristics. Clearly, nonresponse adjusted weights could not be used in the comparison of respondents and nonrespondents, due to the lack of nonresponse adjusted weights for the latter group.

[3] A child was considered a respondent in the base year if he had a completed child assessment or parent interview in the fall or spring of kindergarten.

There were statistically significant differences[4] between parent interview respondents and nonrespondents that were consistent across the fall and spring first grade. A higher percent of children with parent interview data in each rounds were White, food secure, attending private schools, and were from two-parent, high socioeconomic, non-poverty, English-speaking households with higher maternal education. Table 3 illustrates some of the differences in child and household characteristics. Interestingly, these differences were not as consistently prevalent when we examined children with and without a completed direct child assessment in rounds 3 and 4. The main difference between respondents and nonrespondents to the direct child assessments in both rounds was that a higher percent of respondents came from two-parent families.

Why would there be such strong differences between children with and without a completed parent interview, but not between children with and without direct child assessment data? One possible explanation is that the reasons for non-participation differ between the parent interviews and child assessments. Many of parents directly refused to participate, were difficult to get a hold of, or were unable to complete the parent interview, e.g., due to language restrictions. However, in the case of the children, very few children directly refused to participate in the assessment, and in addition, in first grade only a few of the parents who refused to complete a parent interview stopped their children from participating in the child assessment. In most cases, the main source of nonresponse for the child assessments was the child's absence on the days that ECLS-K staff was in their school. This difference in the reasons and magnitude of nonresponse is reflected in the differences between the nonrespondents to the two different instruments.

Thus, the answer to our first question on whether there are differences between respondents and nonrespondents to the parent interview and child assessments, would be that there are. So we explored our next question: do these differences contribute to potential bias? By this we mean: if we were to derive an estimate of a base year characteristic (using data collected in the base year) derived from the first grade respondents would the estimate differ from one derived from all the base year respondents? (The base year respondents include both the first grade respondents and nonrespondents.)

If an estimate of a base year characteristic for the combined sample of first grade respondents and nonrespondents is different from the estimate based on just the first grade respondents, then this difference can be quantified as nonresponse bias. Thus, the next step was to compare base-year estimates using the entire sample to estimates using just the first grade respondents for those variables identified as potentially contributing to bias in the previous step where we compared first

[4] All differences are significant at the 0.05 level.

grade respondents and nonrespondents. For example, 59 percent of the respondents were White, whereas 41 percent of the nonrespondents were White. Thus, if a researcher estimated the proportion of White kindergartners using just the spring first grade parent interview respondents, he would obtain an estimate that 59 percent of the kindergartners were White. However, the same estimate using both the first grade respondents and nonrespondents would be 56 percent. If these two estimates were significantly different then this difference could be considered as a quantification of bias prior to using nonresponse-adjusted weights in the estimation process.

Of the 200 comparisons, 160 were not statistically significant. Of the remaining 40 statistically significant differences, there were no consistent patterns across the different instruments or rounds of data collection. For example, the estimate of the percent White kindergartners was statistically different based on whether we used spring first grade parent interview respondents or the entire sample. However, the estimate of White kindergartners was not statistically different if we used any of the three other respondent groups (fall first grade parent interview, fall first grade child assessment, and spring first grade child assessment respondents) and compared those estimates to the estimate from the overall sample.

These remaining differences lead us to our third question: Even though there are some *statistically* significant differences, how *substantively* different are they? In most of the cases, the differences are neither consistent across different groups of respondents, nor are they substantively significant. For example, 21 percent of the children with completed round 4 parent interview data had mothers with a bachelor's degree or higher compared to 20 percent of the children in the entire sample. Similarly, the mean spring kindergarten math scale score for spring first grade child assessment respondents was 26.9 compared to 26.6 for the entire sample. The standard deviation for the spring kindergarten math scales scores is 8.8; thus a difference of 0.3 points represents an effect size of only 3.4 percent.

Even through there were differences between respondents and nonrespondents in personal and familial characteristics, as well as in assessment scores, overall there were very few cases of potential bias. Part of this is due to the relatively high completion rates. In the cases were there was statistically significant bias, the magnitude of the bias was very small. In surveys with large sample sizes like the ECLS-K, often differences in estimates are statistically significant due to the large sample sizes, but not substantively important.

6. Simulating nonresponse

Since the ECLS-K is a longitudinal study that has multiple components (e.g., child assessment, parent interview, teacher questionnaires) and these components have different response rates, different weights have been created for use in various analyses. Several different cross-sectional and longitudinal weights were developed depending on which of the components were completed within and across waves for the children. For example, for the spring first grade (round 4) three cross-sectional weights were created: one based on the presence of child assessment data (C4CW0), a second based on the presence of parent interview data (C4PW0) and a third based on the presence of child assessment, parent interview *and* teacher questionnaire data (C4CPTW0). Thus, the children with a positive C4CPTW0 weight are a subset of the children with a positive C4CW0 weight. The difference between these two groups of children is due to additional nonresponse in the parent and teacher components. If a child had a completed direct child assessment, but no parent or teacher data, then he would have a positive C4CW0 weight and a zero C4CPTW0 weight.

Similar series exist longitudinally, where the weights are based on increasing levels on longitudinal nonresponse. For example, we have weights based on the presence of child assessment data from round 1 (C1CW0), rounds 1 and 2 (BYCW0), and rounds 1, 2 and 4 (C124CW0). Each progressive weight is defined by increasing levels of nonresponse, e.g., a child who has just round 1 child data won't have a positive BYCW0 and C124CW0 weight, and a child who just has rounds 1 and 2 child data won't have a positive C124CW0 weight. Children with a positive C124CW0 weight belong to the smallest subgroup: they have all three rounds of child assessment data. Thus, if these three weights were used to generate an estimate, any difference in the estimates could be attributed to the effects of increasing levels of nonresponse remaining even after the nonresponse adjustments.

The variability in the number of completed components for the children provides a realistic basis for simulating nonresponse. Having multiple definitions of response provides the opportunity to estimate the same characteristic using different sets of respondents and the corresponding nonresponse adjusted weights. The differences between the estimates are only due to 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 4, the difference between an estimate from the child assessment using the C4CW0 weight (based on all the children with these data) and the C4CPTW0 weight (the smallest subset of responding children with these data) measures the differential nonresponse bias associated with the observed response pattern.

Direct child assessment scores, psychomotor skills, and parent and teacher ratings of children's socioemotional status were estimated for both longitudinal weights based

on four waves of data collection and cross-sectional weights based separately on fall and spring first grade data. There were no systematic substantive differences in scores depending on the different weights used. This suggests that either nonresponse bias was not introduced despite the increasing levels of nonresponse, or that the nonresponse adjustments for these weights accounted for any bias that might have been introduced. Tables 4 and 5 illustrate estimates based on a cross-sectional and longitudinal weight series.

7. Conclusions

This effort is a continuation of the nonresponse bias analysis work started with the first two rounds of the ECLS-K. The base year work allowed us to target potential sources of nonresponse bias and to decrease the risk of bias in survey estimates for later rounds. Due to the identification of low school administrator questionnaire response rates in high minority schools, we were able to target these schools for intensive nonresponse follow up in spring first grade and improve the response rates, despite the overall decreases in response rates.

Since bias is considered to have two components, the extent of nonresponse and the difference between respondents and nonrespondents, we also looked closer at the characteristics of respondents and nonrespondents to the different survey instruments. In a longitudinal study, an advantage in subsequent rounds is that we have rich information on the later round respondents and nonrespondents, as they had been respondents during earlier waves. This wealth of information helps us understand the effects of attrition in subsequent rounds.

We examined completion rates, differences between respondents and nonrespondents to various instruments, and differences in estimates constructed using weights with increasing levels of nonresponse. The completion rate patterns among the various subgroups looked very similar to the base year patterns, and we were able to improve on previously low subgroup rates. While there were differences between respondents and nonrespondents, and a few of these differences were statistically significant, there were virtually no differences that would be considered substantively significant. And finally, estimates using weights based on increasing levels of nonresponse were very similar, thus implying a lack of bias due to nonresponse, or remaining after the nonresponse adjustments to the weights.

8. Limitations

One limitation of our study is that the characteristics we are using to describe the first grade respondents and nonrespondents are from one time point—base year, and may have changed in the first grade. A second limitation is that the study examines additional bias introduced into the sample through attrition, using variables that were available from the base year respondents. It does not evaluate the original sample and the final sample in terms of bias introduced in the survey due to nonresponse across the four waves of data collection. A bias analysis was conducted after the first two waves of data collection, where no strong sources of potential nonresponse bias were identified. The assumption of this effort is that, if no additional nonresponse bias is identified after rounds 3 and 4, then overall, across the four rounds there should be minimal nonresponse bias in the data. Also, this study primarily studies the effect of unit nonresponse. Item nonresponse can be an additional source of nonresponse bias.

9. 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.

Brick, J., and Bose, J. (2001). Analysis of Potential Nonresponse Bias. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, (CD ROM).

National Center for Education Statistics. (2001). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99: Base Year Public-Use User's Manual.* NCES 2001-029. Washington, DC: U.S. Government Printing Office.

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

National Center for Education Statistics. (2002b). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99: First Grade Public-Use User's Manual.* NCES 2002-135. Washington, DC: U.S. Government Printing Office.

Table 1.— Weighted completion rates for the ECLS-K spring first grade school administrator questionnaire by percent minority enrollment

Percent Minority Enrollment	Spring First Grade Completion Rate (%)	Change in Completion Rate from Base Year (%)	Number of Schools (count)
0-10 percent	87	-6	4,879
11-49 percent	84	-9	4,729
50-89 percent	79	0	2,508
90-100 percent	77	15	2,425

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, User's Manual for the Base Year Public-Use Data Files (NCES 2001-029r)

Table 2.— Weighted child-level completion rates for ECLS-K spring-first grade, by child's mover status

Mover Status	Child assessment (%)	Parent interview (%)	School Administrator Questionnaire (%)
All children	88	85	76
Mover status			
Mover	63	74	35
Located, followed	86	78	49
Other*	0	60	0
Nonmover	96	88	89

* This category includes movers who could not be located, movers whose cases could not be processed before the end of the field period, and movers who moved into nonsampled PSUs.

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, User's Manual for the Base Year Public-Use Data Files (NCES 2001-029r)

Table 3.— Characteristics of children with (respondents) and without (nonrespondents) ECLS-K spring first grade parent interview data

Characteristics	Respondents (%)	Nonrespondents (%)
Child's Race— White	59	41
Two-Parent Family	75	61
Food Secure Household	91	87
Non-English Speaking Household	11	18
Mother's Education— BA or Higher	21	11
Household in Top 20 % of SES	19	10

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Unreleased Methodological Variables, First Grade Data Files

Table 4.— Mean of fall kindergarten reading, mathematics and general knowledge assessment scores by a longitudinal series of weights

Fall Kindergarten Assessment	C1CW0 weight	BYCW0 weight	C124CW0 weight
Reading Scale Score	20.39	20.41	20.37
Mathematics Scale Score	18.86	18.86	18.86
General Knowledge Scale Score	20.34	20.39	20.32

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, Base Year Public-Use Data Files (NCES 2001-029r)

Table 5.— Mean spring first grade reading, mathematics and general knowledge assessment scores by round 4 cross-sectional weights

Spring First Grade Assessment	C4CW0 weight	C4CPTW0 weight
Reading Scale Score	54.77	55.38
Mathematics Scale Score	42.79	43.11
General Knowledge Scale Score	34.00	34.25

Source: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, First Grade Public-Use Data Files (NCES 2002-134)