DOCUMENT RESUME

ED 312 311 TM 014 109

AUTHOR	1ngo. s, Steven J.; And Others
TITLE	Schoo ., Individual and Item Nonresponse in the National Education Longitudinal Study of 1988 (NELS:88) Base Year Survey.
INSTITUTION	National Opinion Research Center, Chicago, Ill.
SPONS AGENCY	National Center for Education Statistics (ED), Washington, DC.
PUB DATE	Mar 89
CONTRACT	NCES-300-86-0010
NOTE	64p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Data Collection; Educational Assessment; Elementary Secondary Education; Estimation (Mathematics); Evaluation Methods; Grade 8; Longitudinal Studies; Middle Schools; *National Surveys; *Questionnaires; Research Methodology; Response Rates (Questionnaires); *Sampling; *School Surveys; *Testing Problems
IDENTIFIERS	High School and Beyond (NCES); *National Education Longitudinal Study 1988; *Nonresponders; Student Surveys

ABSTRACT

Nonresponse issues are investigated for the base year (1988) survey of the United States Department of Education's National Education Longitudinal Study of 1988 (NELS:88), a national probability sample of middle schools and eighth-grade students in the spring of 1988. The total eighth-grade enrollment for the NELS:88 sample of schools was 203,002; of these, 10,583 students were excluded due to limitations in their language proficiency or to mental or physical disabilities. As in the 1980 High School and Beyond (HSB) Study, the NELS:88 sample included about 70% initial selections and 30% replacement schools. NELS:88 non-respondents were 51.96% male and 43.07% female, with the gender of 4.96% not indicated, while NELS:88 respondents were 49.8% male and 50.2% female. Part 1 of this paper gives a brief overview of the study and its sample design, and outlines the main non-response issues, namely: (1) school and individual ineligibility for the study; (2) unit non-response, i.e., the fact that some schools and individuals declined to participate; and (3) item non-response in the student questionnaires and cognitive tests. Part 2 describes the methodology used for adjusting school level non-response in the NELS:88. Actual estimators are given, along with a method for evaluating the estimators and a method for deriving the estimate of response propensities for each school. Part 3 reports the results of an item non-response analysis of the student questionnaire data and the cognitive tests. Some comparisons are offered with non-response in the first year of the HSB survey. (SLD)

SCHOOL, INDIVIDUAL AND ITEM NONRESPONSE IN THE NATIONAL EDUCATION LONGITUDINAL STUDY OF 1988 (NELS:88) BASE YEAR SURVEY

STEVEN J. INGELS
LOUIS RIZZO
KENNETH A. RASINSKI

National Opinion Research Center
University of Chicago

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California, March 29, 1989.

The study upon which this paper is based was sponsored by the National Center for Education Statistics (NCES), the Office of Educational Research and Improvement, United States Department of Education. Dr. Jeffrey A. Owings was the NCES Project Officer for the NELS:88 Base Year. The ideas and opinions expressed in this paper are those of the authors and do not necessarily represent positions or conclusions endorsed by NCES.

NORC Projects 4455-4456; NCES Contract 300-86-0010

SCHOOL, INDIVIDUAL AND ITEM NONRESPONSE IN THE NATIONAL EDUCATION LONGITUDINAL STUDY OF 1988 (NELS:88) BASE YEAR SURVEY

by STEVEN J. INGELS, LOUIS RIZZO, KENNETH A. RASINSKI
NORC/University of Chicago

Abstract.

This paper deals with nonresponse issues in the base year (1988) survey of the U.S. Department of Education's National Education Longitudinal Study of 1988 (NELS:88), a national probability sample of middle schools and eighth grade students in the spring of 1988. There are three main parts to the paper.

In Part I, Ingels gives a brief overview of the study and its sample design, and outlines the main nonresponse issues. These issues are:

- School and individual ineligibility for the study (which schools and students were excluded, and the possible implications of these exclusions for national estimation using NELS:88 data);
- Unit nonresponse: the fact that both some selected schools and some individuals declined to participate; the use of nonresponse adjustments in the weighting as partial compensation for possible nonresponse bias;
- Item nonresponse in the student questionnaire and cognitive tests: given that missing data were not imputed in NELS:88, the scope and implications of item nonresponse and attendant biases must also be assessed.

In Part II, Rizzo describes the methodology used for adjusting for school level nonresponse in NELS:88. First he gives the actual estimators of the finite population means adjusted for nonresponse. Then he presents a method for evaluating these estimators. Finally, he describes the methodology for deriving the critical variable for calculating adjusted estimates of population means--the estimate of response propensities for each school in the sample.

In Part III, Ingels and Rasinski report the results of an item nonresponse analysis that they conducted on the student questionnaire data and summarize nonresponse data for the cognitive tests. First, they quantify overall nonresponse and nonresponse bias for key variables on the student questionnaire. Second, they describe nonresponse patterns, both in terms of characteristics of items (content, format, position in the question [early, middle, late]) and of respondents (demographic characteristics of item nonrespondents). Finally, they offer basic nonresponse statistics for the cognitive test battery and provide some comparisons of NELS:88 Base Year student nonresponse with nonresponse in the Base Year of the earlier (1980) High School and Beyond survey.

SCHOOL, INDIVIDUAL AND ITEM NONRESPONSE IN THE NATIONAL EDUCATION LONGITUDINAL STUDY OF 1988 (NELS:88) BASE YEAR SURVEY

Steven J. Ingels, Louis Rizzo, Kenneth A. Rasinski
NORC
A Social Science Research Center
University of Chicago
1155 E. 60th Street
Chicago, Illinois 60637

American Educational Research Association Annual Meetings,
San Francisco, California, March 29, 1989, Session 26.19:
A First Glimpse at NELS:88: Symposium on Design and Findings.

CONTENTS

List of Tables, Figures, and Appendices.

TABLES.

FIGURES.

REFERENCES.

APPENDICES.

AERA, San Francisco
March 29, 1989

I. NELS:88 Base Year Nonresponse Issues - Steven Ingels

In Part I of the this paper, we provide an overview of NELS:88 and its base year sample design, then summarize the incipal unit and item nonresponse issues that portend in this survey.

I-A. OVERVIEW OF NELS:88 and BASE YEAR SAMPLE DESIGN

Overview.

This longitudinal study, by beginning with a cross-section of approximately 26,000 1988 eighth graders, following a substantial subsample of these students in 1990 and thereafter, and by freshening* the sample to ensure tenth and twelfth grade nationally representative student samples in 1990 and 1992, will provide a point of comparison with the high school classes of 1980 and 1982, studied by High School and Beyond (HS&B), and the high school class of 1972, studied by the National Longitudinal Study of the Class of 1972 (NLS-72).

The overall scope and longitudinal design of the study offer at least five major benefits:

(1) Longitudinal focus. First, the study provides the basis for within-cohort comparison by following the same individuals over time. NELS:88 will thus provide measures not only of educational attainment but also explanations of the reasons for and consequences of academic success and failure. (Coincidentally, NELS:88 data will permit the most comprehensive and sophisticated assessment to date of the cumulative impact of the recent school reform movement.) Individual and group level change is captured by NELS:88 in particular by its emphasis on the measurement of cognitive growth and the recording of key transitions.

Many if not most of the questions policy makers seek to answer involve some notion of change over time. Although cross-sectional analysis may approximate the study of the process of change by using a number of devices, the risk involved is substantial and the possibility of examining causal relationships is nil. For example, a study may ask respondents who have failed to complete their elementary education when they started to think about or plan on dropping out, but the danger here is that the farther back they are forced to reach into their memories, the less accurate they are likely to be. (On the unreliability and biases of retrospective survey responses, see especially Bradburn, Rips and Shevell, 1987; and Schuman and Kalton, 1986). Even apart from the limitations and distortions of retrospective accounts, however, cross-sectional approaches are not suitable vehicles for measuring individual change nor do they provide a viable basis for causal inference.

(2) Representative national cross-section. The second benefit of the NELS:88 design is that it provides a representative cross-section of eighth graders in the United States, thereby shedding light on the factors at the demographic and environmental levels that affect educational outcomes such as school performance and individual aspirations. Thus--and unlike its predecessor national longitudinal studies, NLS-72 and HS&B, NELS:88 establishes a baseline for measuring the impact of secondary schooling at a point just prior to entry into high school.

*For an explanation of sample freshening in NELS:88, see Appendix 2.

(3) Transitions of early adolescence. The third advantage of the NELS:88 design is that it begins with early adolescence, a developmental transition period in which major changes in individual attitudes and behaviors take place, and a time point that also marks the transition to secondary schooling. Thus it provides a basis for understanding such areas of concern as the impact of tracking and school and program choice in the middle years on subsequent educational and occupational outcomes, and the interaction between schooling and the crystallization of key attitudes, values and aspirations. While earlier studies such as NLS-72 and HS&B monitored the critical transition from high school completion to the labor force, postsecondary education, and family formation, the singular strength of the NELS:88 design is that it also encompasses the key school transition from eighth grade to high school and associated developmental transitions of early adolescence.

(4) Trend analyses. The fourth design benefit is that NELS:88 offers the opportunity for the analysis of trends in areas such as academic performance. Cross-cohort comparisons with earlier NCES longitudinal studies will be possible as early as the 1990 wave of data collection.

(5) Holistic perspective. NELS:88 takes the student as the fundamental unit of analysis, further illuminating the student data by tapping the rich contextual information available from other respondent populations and records sources. NELS:88's major features--the planned integration of school administrator and records, student, parent and teacher studies and the inclusion of supplementary components to support analyses of demographically distinct subgroups such as Hispanics and Asians, stamp the study with an exceptionally comprehensive research design.

The data produced through this design can facilitate the development and evaluation of educational policy at all governmental levels. The NELS:88 data can also inform decision-makers, educational practitioners, and parents about the changes in the operation of the educational system across time, and the effects of various elements of the system on the lives of the individuals who pass through it. Thus the base year and follow-up studies explore a number of areas that define the basic outcome variables of NELS:88--those related to cognitive growth, occupational expectations and achievement, and personal and social development. Information has been gathered as well on numerous independent variables, such as standard demographics, and variables measuring educational support, parent's socioeconomic status, family composition, language use, and home environment. The core of intervening variables encompasses school experiences such as exposure to given curriculum content and structure, assessment and evaluation systems, social relations, school behavior, and participation in extracurricular activities.

Sample Design.

Selection of Schools and Students

The NELS:88 Base Year sample is representative of eligible eighth grade schools and students in the fifty states and the District of Columbia. NELS:88 students were sampled through a two-stage process, modelled after that used for the two prior NCES longitudinal surveys, NLS-72 and HS&B.

The first stage involved stratified sampling of over 1,000 public and private schools from a universe of approximately 40,000 schools containing eighth grade students. Stratification is by administrative control, with an oversample of Catholic and other private schools; and by geography, permitting comparisons among the nine Census divisions. The sample was drawn with probabilities proportionate to a school's eighth grade enrollment.

The second stage included random selection of approximately 26 (24 core, and, on average, 2.2 oversampled Hispanic and Asian supplemental) students per school. Once students are selected, the NELS:88 sample design includes one parent for each selected student, and two teachers in designated subject areas. Follow-up surveys are planned for 1990 and 1992, when most of the initial cohort will be in the tenth (1990) and twelfth (1992) grades, and 1994. Students who drop out of school during this period will also be surveyed.

NELS:88 eighth graders completed a questionnaire, and a series of four tests (in reading, mathematics, social studies [=history/citizenship] and science), that are designed to measure cognitive growth over time. (For a comprehensive account of the development and psychometric characteristics of the NELS:88 cognitive test battery, see Rock and Pollack, in Ingels et al., 1987).

The Parent Sample

One parent of each child has been included in the study. Parent data will be used primarily in the analysis of student behaviors and outcomes, and only secondarily as a dataset by itself. Parents completed a self-administered questionnaire that sought information on home background and education support system and the family's interactions with the school.

Teacher and Administrator Samples

All full- and part-time instructors who are teaching classes in mathematics, science, English/language arts, and social studies to eighth graders in the spring of 1988 were included in the NELS:88 universe of eighth grade teachers. The actual sample was restricted to teachers who provided instruction in the listed subjects to the selected sample of eighth grade students within the sampled schools. Two teachers were sampled for each selected student. The administrative head of each school was also included in the sample. Again, since the student is the unit of analysis, the NELS:88 sample is not a national probability sample of eighth grade teachers, but of selected teachers of a nationally representative sample of eighth graders. The teacher questionnaire gathered contextual data on individual students, classes, and the teacher and school. The school administrator questionnaire provided additional school context data.

The Asian and Hispanic Supplemental Samples

The NELS:88 design consists of a core sample, plus a supplementary sample of students (and parents and teachers) sponsored by the Office of Bilingual Education and Minority Language Affairs (OBEMLA). The OBEMLA supplement includes additional students of Hispanic and Asian descent beyond the numbers selected for the core study. This oversample of Hispanic and Asian students supplied analytically sufficient numbers of these groups and their principal subgroups (for example, Hispanic Americans of Mexican, Puerto Rican, Cuban, and other Hispanic descent; Asian students of Chinese, Japanese, Filipino, Korean or Southeast Asian descent) and statuses (for example, recent immigrants, versus Asians and Hispanics long established in the United States; bilingual "language minority" students, and students with moderately limited English proficiency, as well as Asian and Hispanic English monolinguals). Nevertheless, the NELS:88 cognitive tests and questionnaires were available only in English. Students with no or with severely limited English proficiency were therefore excluded from the sample. This exclusion qualifies the representativeness of the NELS:88 Hispanic and Asian student samples.

A more detailed account of the NELS:88 sample design can be found in the forthcoming NCES publication, NELS:88 Base Year Sample Design Report (Spencer, Frankel, Ingels, Rasinski and Tourangeau), while a description of the NELS:88 data files is to be found in NELS:88 Base Year User's Manuals (Ingels, Abraham, Carr, Frankel, Rasinski, Spencer), forthcoming from NCES late in 1989.

I-B. NELS:88 NONRESPONSE ISSUES: A SUMMARY

Two distinct kinds of unit nonparticipation can threaten the integrity of a national probability sample of schools and students such as the NELS:88 Base Year.

(1) At one level, there is the issue of eligibility--are certain schools, or students, excluded from participation?; if so, does this exclusion undermine the capacity of the study's data to produce national estimates and sustain valid analyses?

(2) A second unit participation issue is school and student nonresponse: since, both at the school level and at the individual level, populations of respondents and nonrespondents may significantly differ in their characteristics, the sample statistics may be biased as estimates of the characteristics of the entire population.

Before examining eligibility and nonresponse in detail, it may be useful to summarize the various ways in which a student may fail to have a chance of selection into the dataset. Basically there are seven such ways:

(a) First, if the student's school refused, that student had no chance of selection;

(b) Second, if the student's school was declared ineligible to participate, that student had no chance of selection;

(c) Third, though the selected school participated, the student was declared ineligible to participate, owing to physical or mental handicaps, behavioral problems, or a lack of command of English;

(d) Fourth, the student was studying at home in 1987-88, or abroad;

(e) Fifth, the student was temporarily unavailable (for example, was hospitalized during the survey period, or was a migrant in transit);

(f) Sixth, owing to clerical error, the student did not appear on the correct roster or was misclassified. (While we believe that in general school rosters were extremely accurate, there is some evidence that transfer-ins between the time of initial sampling and the sample update just before Survey Day were, as a group, sometimes missed);

(g) Seventh, the student's school had no chance of selection, because the sampling frame was inaccurate (for example, a student might attend a newly-opened school that had not yet been added to the school list from which the sample was drawn).

Even if the student's school participated and the student was declared eligible and is selected, there are two additional ways that the student's data might fail to be included in the dataset:

(a) First, though selected, the student refused to participate; or

(b) Second, the selected student participated at some level, but did not fulfill the restrictive participation condition of having student questionnaire data--either because the student did not complete the questionnaire (though the student may have completed the test); or because of the occurrence of rare events such as loss of the questionnaire in transit or its accidental destruction in the optical scanning process.

We now turn to the issue of ineligibility--what schools and what students were systematically excluded from participation in NELS:88, and what are the implications of their exclusion? After we have addressed questions about eligibility and its implications, we shall go on to examine the matters of school and individual nonresponse.

Ineligibility: Excluded students, excluded schools.

The kinds of people excluded from the target population fall into three main classes, specifically, mentally handicapped persons and students not proficient in English for whom the NELS:88 tests would be unsuitable and persons having physical or emotional problems that would make participation in the survey unduly difficult or unwise. To enable inferences to be extended to the larger populations which include these excluded persons, we collected data on the numbers of students excluded as a result of these restrictions.

Seven ineligibility codes were employed at the time of student sample selection. (These criteria are similar but not identical to the HS&B codes: see Exhibit B in Appendix 1). In NELS:88, a student was ineligible to participate if that student:

A - only attends the sampled school on a part-time basis, primary enrollment at another school

B - physical disability precludes filling out questionnaires and tests

C - mental disability precludes filling out questionnaires and tests

D - dropout: absent or truant for 20 consecutive days, and is not expected to return to school

E - does not have English as the mother tongue AND he or she has insufficient command of English to complete the NELS:88 questionnaires and tests

F - has transferred out of the school since roster was compiled

G - is deceased

In cases D, F, and G, the student was no longer at the school. In cases A, B, C, and E, the student, though still enrolled at the school, was excluded from the sample. This special subset of the ineligibles--excluded students--is of special interest, since three of the four exclusion categories have implications for national population estimates projected from the NELS:88 dataset.

Part-time status (code A) has no implications for estimation. These students had their primary enrollment at another school. (Each eighth grade student was to have one <u>and only one</u> first-stage [that is, school-level] chance of selection into the NELS:88 sample). However, exclusion of cases covered by codes B, C, and E — ineligibility keyed to physical, mental or linguistic difficulties in completing the instruments -- has implications for estimates drawn from the Base Year sample and subsequent study waves. It may therefore be useful to say more about how exclusion criteria were applied, and what numbers of students were excluded for each of these three categories.

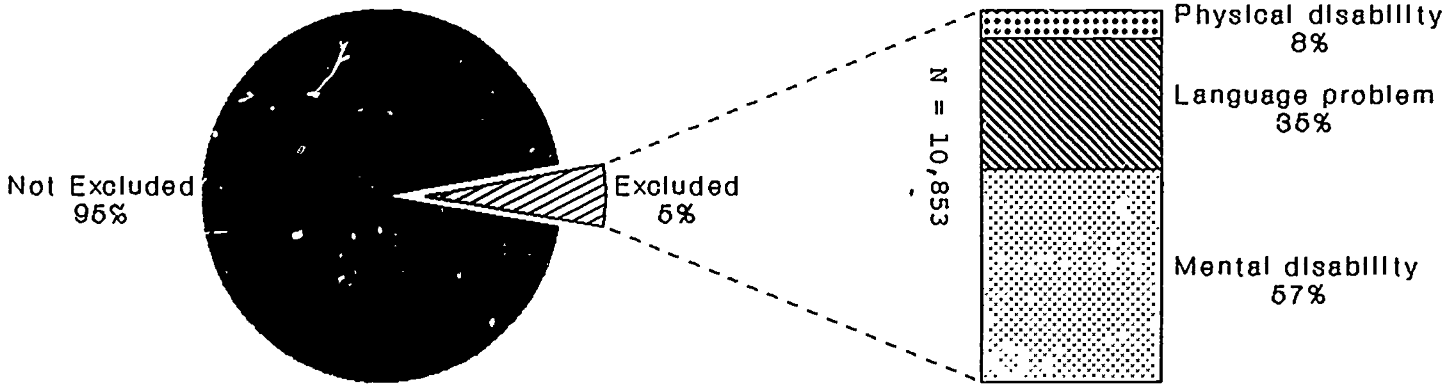
School Coordinators were asked to indicate which students should be excluded, and annotated each excluded student's entry on the sampling roster with the appropriate code. Exclusion decisions were to be made on an individual basis. Thus special education and Limited English Proficiency (LEP) students were not to be excluded categorically. Rather, each student's case should be reviewed to determine the extent of limitation in relation to the prospect for meaningful survey participation. Students, including LEPs and physically or mentally handicapped students were to be surveyed if school staff deemed them capable of completing the NELS:88 instruments, and excluded if school staff judged them, on an individual basis, to be incapable of doing so. School coordinators were told that when there was doubt, they should include the student.

The total eighth grade enrollment for the NELS:88 sample of schools was 203,002. Of these 203,002 students, 10,853 were excluded owing to limitations in their language proficiency or to mental or physical disabilities. Thus 5.34 percent of the potential student sample (the students enrolled in grade 8 in the selected NELS:88 schools) were excluded. Breakdowns by exclusion categories are as follows (and see Figures I-1 and I-2):

Figure I-1

Excluded Students as a Proportion of the Potential Student Sample

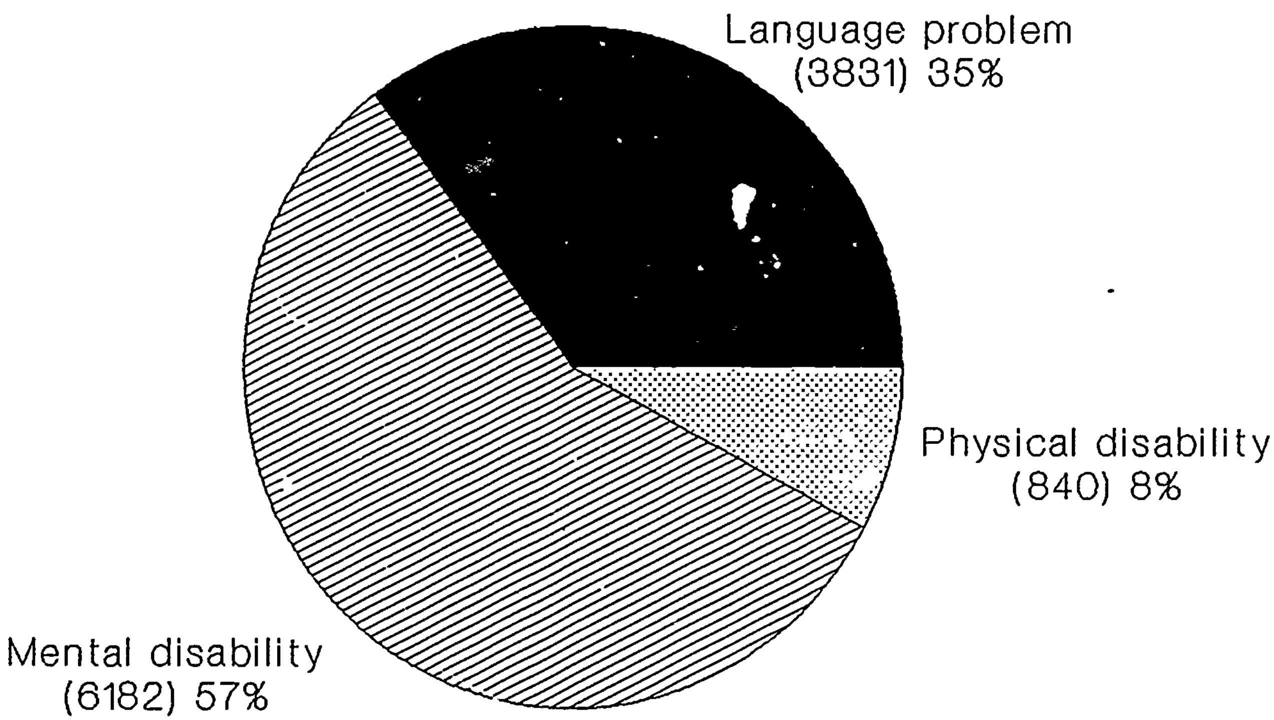
NELS:88 Base Year Schools



N=201,917 (Total Eighth Grade Population of NELS:88 Base Year Schools)

Figure I-2

Excluded Students
NELS:88 Base Year Schools



N=10,853

Ingels, Rizzo, Rasinski
AERA -- March 1989

840 -- B - excluded: PHYSICAL DISABILITY

6,182 -- C - excluded: MENTAL DISABILITY

3,831 -- E - excluded: TOO LIMITED PROFICIENCY IN ENGLISH

Less than one half of one percent of the potential sample was excluded for physical disability (.41 percent), but 3.04 percent was excluded for reasons of mental disability, and 1.88 percent because of limitations in English proficiency. Put another way, of the 10,853 excluded students, 57 percent were excluded for mental disability, 35 percent owing to language problems, and 8 percent because of physical disabilities. Since current characteristics and probable future educational outcomes for these groups depart in many ways from the national norm, the exclusion factor should be taken into consideration in generalizing from the NELS:88 sample to eighth graders in the nation as a whole.

An overall exclusion of 5.4 percent can of course translate into a much higher rate in some localities with high immigrant populations (New York City and Los Angeles are examples of the extreme cases; in the New York City public schools, for example, ineligibility rates exceeded 20 percent).

This implication for estimation carries to future waves. For example, if the overall propensity to drop out is twice as high for excluded students as for non-excluded students, the dropout figures derivable from the NELS:88 First Follow-Up study would underestimate early (1990) dropouts by around ten percent.

Just as certain students were considered to be ineligible, so too certain kinds of schools were ineligible for selection. The eligible populations of schools are restricted to "regular" schools in the U.S., private as well as public. Excluded from the sample are Bureau of Indian Affairs (BIA) schools, special education schools for the handicapped, area vocational schools that do not enroll students directly, and schools for dependents of U.S. personnel overseas.

The impact of school ineligibility on the representativeness of the NELS:88 sample or the generalizability of its data should be quite small. Most of the special education schools for the severely handicapped are not age-graded and thus do not contain eighth grades as such. Less than 10 percent of the American Indian population is enrolled in BIA or BIA contract schools (of the more than 400,000 American Indian students nationwide, 6.4 percent are enrolled in schools directly operated by the BIA and another 2.9 percent in BIA schools operated by Indian communities). The number of students enrolled in Department of Defense Dependents Schools in 1988 in the eighth grade was fewer than 10,000. (In contrast, the number of eighth graders enrolled in schools eligible for NELS:88 exceeded three million.) It should be noted that some students excluded in 1988 by virtue of attending ineligible schools will, under certain circumstances, have some chance of selection into the NELS:88 sample in the 1990 First Follow-Up. Thus a 1988 BIA student who transfers to a non-BIA school in 1990 has a theoretic possibility of being "freshened" into the NELS:88 sample. 1988 eighth graders then enrolled in overseas Department of Defense Dependent Schools who are repatriated between spring of 1988 and autumn of 1989 and are enrolled in tenth grade during the 1989-90 school year likewise are conferred a chance of selection [through sample "freshening"] into the First Follow-Up.

Of course, students who are educated at home or in private tutorial settings, and those who have dropped out of school prior to reaching eighth grade also fall outside the NELS:88 Base Year sample. These exclusions too have implications for national inferences based on NELS:88 data, although as in the case of school ineligibility, their impact on national estimates generally should be quite small. The one sample coverage and eligibility issue, then, that may have a significant impact on estimation, is that of excluded students. In order to address this problem, NORC and NCES have discussed the desirability of following a subsample of NELS:88 Base Year ineligibles in the 1990 First Follow-Up and again in 1992. Basic demographic information would be obtained about this group, and an event history of their school status would be constructed. In this way a correction factor for national estimation from NELS:88 data of, for example, dropout rates, could be formulated.

Students who are excluded are excluded because it is extremely difficult if not impossible to test and survey them. However, the fact that they cannot be surveyed does not mean that additional data cannot be collected about them (NAEP, for example, has school officials complete an excluded student questionnaire). In addition, it is important to ensure that all students who in fact can be surveyed and tested are included in the survey. HS&B, NELS:88 and NAEP have all worked to reach this goal, and NAEP has recently devised a somewhat more stringent set of rules for defining linguistic ineligibility which, if successful, may reduce NAEP's overall exclusion rate from around 5 percent to 4 or 4.5 percent. (See Exhibit A in Appendix 1 for a statement of these criteria). For future longitudinal studies in the NELS:88 mold, one well might wonder if students in those ineligibility statuses that are potentially transient--one thinks, in particular, of limitations in language proficiency--might not be followed over time, on the assumption that in later waves their command of English well might permit them to participate.

School-level nonresponse.

The problem of school-level nonresponse is a serious one for school-based surveys. To the extent that populations of respondents and nonrespondents differ, the sample statistics will be biased as estimates of the characteristics of the entire population. The magnitude of the bias introduced into means and proportions by nonresponse depends upon two factors--the size of the nonrespondent stratum and the difference between units in the responding and nonresponding strata. In a two-stage (school and student) sampling process, bias can result both from school nonresponse, and from student nonresponse. The effects of the two types of nonresponse are additive.

As in High School and Beyond, the NELS:88 sample is made up of approximately 70 percent initial selections, and 30 percent replacement schools. It is anticipated that in future waves of NELS:88, as in the later school-based wave of High School and Beyond, school cooperation rates will exceed 90 percent--once such a study is begun, schools feel keenly the obligation to honor a request to participate, given that no substitution of students is possible in a longitudinal study. Nonetheless, however high the school cooperation rate in the subsequent rounds of a longitudinal study, any biases attributable to the initial (baseline) sample will persist in the later waves. Students attending schools that did not cooperate in the Base Year were not sampled and have little or no chance of selection into the Follow-Ups. If these students do indeed differ in significant respects from students at cooperating schools, the bias introduced by Base Year school noncooperation will affect all successive waves. (Note, however, that noncooperation at the student level does not have such implications for the subsequent rounds of a longitudinal study because Base Year nonrespondents remain in the sample and can be surveyed in the future.)

The Base Year replacement strategy was designed to make substitution schools as similar as possible in key characteristics to the school which they replaced--thus for example, a suburban, Catholic school in a given geographic superstratum was always replaced by another suburban Catholic school within the same geographic superstratum. Nevertheless, there are many other school characteristics in which a refusing initial selection and its replacement may differ. Moreover, a key assumption here is that outwardly similar schools will contain essentially similar students. While this is a reasonable assumption, the degree to which it is true is uncertain.

Certain nonresponse adjustments, however, are possible that partially compensate for nonresponse biases. In particular, the bias associated with unit nonresponse at both the school and the individual level can be controlled by making adjustments to case weights, thus correcting errors in the relative frequencies given to various subgroups. For example, if suburban Catholic schools have a higher response rate than urban or rural Catholic schools, the overrepresentation of these schools can be corrected for. However, nonresponse adjustments to the weights do not compensate for another kind of bias--nonresponse bias within subgroups, as when participating suburban Catholic schools are significantly different in other key characteristics from nonparticipating suburban Catholic schools.

Lou Rizzo, in Part II of this paper, describes the methodology for adjusting for school level non-response in the NELS:88 Base Year. Again, however, it must be stressed that the weighting procedures compensate only for one kind of bias, and not all bias. Therefore, one should also compute estimates of the bias still remaining after the weighting adjustments. We have not taken this additional step at this time, but will be doing so as we further document the Base Year sample over coming weeks. The fact that NELS:88 collected school-level data from virtually all participating schools (the school administrator questionnaire weighted response rate was 98.9 percent), and that over 97 percent of the principals of initial school selections that refused to participate in NELS:88 also supplied critical comparison data on the characteristics of their schools, greatly facilitated precise adjustment for nonresponse bias. The presence of such complete survey data for nonrespondents also provides a sturdy basis for estimating school nonresponse bias.

Individual-level nonresponse.

Base Year nonrespondents remain in the NELS:88 sample. A good many of the Base Year nonrespondents will participate in the NELS:88 First Follow-Up in 1990; at that time, we shall get a better reading on their characteristics, although even now at least two basic facts are known about them (sex; and ethnicity, that is, Asian/Hispanic/Neither Asian nor Hispanic) from the school rosters, and additional characteristics (geographical location, school sector) are known from the sampling frame. This being the case, weighting can be used at the individual level too to adjust for differential response rates.

An analysis of participation by (unweighted) sex and ethnicity of Base Year nonrespondents makes clear that Base Year participants and non-participants are indeed not necessarily identical in their characteristics. For example, nonrespondents were 51.96 percent male (NELS:88 respondents were 49.8 percent male), 43.07 percent female (NELS:88 respondents were 50.2 percent female) while for 4.96 percent of nonrespondents gender could not be ascertained (that is, the information was not supplied on the roster and could not be inferred from the student's name). Higher male representation among nonrespondents mirrors HS&B's experience with older adolescents in 1980; also, adult surveys using full probability sampling generally have an underrrepresentation of males. However, it is of interest to see how early this tendency asserts itself, particularly given the fact that at eighth grade the school is still presumed to have vastly more control over the movement and activities of students than would be the case in many secondary schools.

Ethnicity data show that 4.57 percent of non-participating students were Asian, versus 6.2 percent of Base Year participants; 14.61 percent of nonrespondents were Hispanic versus 12.9 percent of respondents; and 80.81 percent of nonrespondents were neither Hispanic nor Asian as were approximately 80 percent of Base Year respondents.

The other major issue that we would address at the individual respondent level is the accuracy of records in identifying the student sample. While overall Base Year sampling procedures worked extremely well, two weaknesses in school records systems were observed that have possible implications for the integrity of the student sample. One such weakness is the ambiguity of dropout status. A second such weakness in the records system is the inability of many schools to provide accurate sample updates--records of all students who have transferred into the eighth grade of the school between the date of generation of the original sampling roster and the date (a week to ten days before the survey session) of the sample update.

Dropouts.

To be classified as a dropout in the NELS:88 Base Year study, several conditions had to be met: the student was absent from school for at least twenty consecutive days, the absence was not excused, and it was the opinion of the school coordinator that the child would not return to school. Since typically one cannot officially leave school until age sixteen (or seventeen, in some locales, such as New York City), schools find it difficult to classify as dropouts most eighth graders who are chronically truant or who are de facto dropouts but have not yet reached the legal school-leaving age. However, given the NELS:88 definition of a dropout, school coordinators reported that 96 sampled students who were absent on survey and makeup days were in fact dropouts--that is, students who had left school, meeting the conditions of the Base Year dropout definition, between the time of original sampling (normally November or December 1987 though sometimes in January or even February of the new year) and the survey session (between February and May of 1988). When NORC traced these students the following autumn, it was found that the dropout status of only twenty-nine could be confirmed. Most of these students had in fact transferred out of their schools. Later requests for school records from the sample school generated status updates that reclassified them as decidedly not school leavers. On the other hand, it is distinctly possible that some number of the students that schools classified as transfers out of the sample in fact had dropped out of school. Reliance on school records to sort out dropouts from transfers is surely a less accurate procedure than following each sample member to determine that student's subsequent status.

Transfer students.

NELS:88 followed essentially the same procedure for dealing with transfer students as did HS&B in 1980. School rosters were submitted and an initial sample drawn in the autumn of 1987 (normally, in November or December, though student samples continued to be drawn through the early months of 1988). To adjust the student sampling frame for student attrition and change in the eighth grade population of the sampled school, we conducted a sample update seven to ten days prior to the school's scheduled survey session. The NORC survey representative went over the sample list with the school coordinator to ensure that all sampled students were still eligible, and that transfer-ins--that is, any student who had joined the eighth grade between the time of original sampling and the time of the update--were added to a supplementary roster from which additional students would be selected. The supplementary roster was annotated for eligibility and ethnicity and the transfer-in students were sequentially numbered. Selections for inclusion in the sample were based on the same set of computer-generated random numbers used to select the original sample and Asian/Hispanic oversamples for that particular school.

In theory, mobile students should be like other mobile students--that is, on an overall national basis, transfer-ins should have essentially the same characteristics as transfer-outs, and there should be a rough parity in the numbers of selected students lost to transfer and the number selected into the sample from the pool of transfers in. However, we recorded a level of outward transfers (975) that was approximately twice that of the number of transfers into the school who were selected into the sample (cf. also the field test results, in Ingels et al. 1987, 66-67). Of course, one is certain to find out that a sample member has transferred--even if that student is missed in the sample update, the student's transfer status will be made apparent on survey day. Schools did not always have accurate and up-to-date lists of transfer-ins, however, particularly when this list needed to take the date of the initial sampling as its baseline. Transfer students are therefore probably somewhat underrepresented in the Base Year dataset and to the extent that transfer students may represent a somewhat different population than non-transfers, this can be a source of some additional individual-level bias.

While the strategy of sampling, then updating the sample, is an intuitively sound one, there are alternatives that are also intellectually attractive and may lessen dependence on school record-keeping and interpretations of ambiguous statuses. One such alternative that might be considered is that of freezing the sample at the time of initial selection. Then all sample members who left the school would be followed, and there would be no attempt to update the sample to give transfer-ins some chance of selection. It may be argued that for many purposes (for example, the analysis of school effects) the sample member who transfers out tells more about the school than does the transfer-in, whose experience really reflects another school. A second advantage would be that in freezing the sample and following all, one overcomes the dropout/transfer ambiguity as well as the transfer in/out numerical asymmetry. However, this would be a somewhat more expensive option than the traditional sample update, because of the expense of following individual students who have left the school, as compared to the economies of group survey sessions within the school.

Item-level nonresponse

Quite apart from the nonresponse of units (whether schools or individuals) item nonresponse--the failure of a respondent to complete certain items on the questionnaire or the test--is a further source of possible bias. While missing data are sometimes imputed, no such procedure has been followed in NELS:88.

In Part III of this paper, Ingels and Rasinski describe and quantify item nonresponse in the NELS:88 Base Year student survey. In analyzing item nonresponse they also identify characteristics that are associated or correlated with the nonrespondents. This information will be valuable in quantifying the maximum and probable bias due to item nonresponse, and in identifying estimates of the populations for which users should use caution.

II. SCHOOL NONRESPONSE ADJUSTMENT IN NELS:88

Louis Rizzo

This paper describes methodology for adjusting for school level non-response for the NELS88 base year survey. The first section gives the actual estimators of the finite population means adjusted for non-response and a methodology for evaluating these estimators. The second section then describes the methodology for estimating propensities of response for each school in the sample - the critical variable for calculating adjusted estimates of population means.

Section 1

The interest here is at the school level, thus the strata mean of interest is

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

where i indicates a particular school i, y_i is a response variable of interest at the school level, and N is the size of a given strata (each strata is treated separately, and aggregated by overall strata weights). Let π_i be the probability school i is selected into the sample. Let s be the set of sampled schools, and n, be the sample size within the strata. Then

$$\hat{\bar{Y}}_{HT} = \frac{1}{N}\sum_{i\varepsilon s} \frac{y_i}{\pi_i}$$

This would be the estimator of $\bar{Y}$, the finite population mean, under full response. Let rs be the responding sample and let $\hat{\rho}_i$ be the probability a given item would respond if selected. The extended Horwitz-Thompson estimator is

$$\bar{Y}_{HT} = \frac{1}{N}\sum_{i\varepsilon r_s} \frac{y_i}{\pi_i \hat{\rho}_i}$$

ρ_i is unknown, and thus needs to be estimated by $\hat{\rho}_i$. The simple approach is to set $\hat{\rho}_i$ as the overall rate of response in the sample. An alternative is to estimate $\hat{\rho}_i$ via logistic regression, which has been done for NELS88. In both cases the estimator of $\bar{Y}$ will be

$$\bar{Y}_{HT}=\frac{1}{N}\sum_{i\varepsilon r_s}\frac{y_i}{\pi_i\,\hat{\rho}_i}$$

The following computations are based on Platek and Gray [1983].

There are three random variables α_i, β_i, and $\hat{\rho}_i$. α_i is defined to be

$$\alpha_i=\begin{cases}1 & \text{if } \textit{item i in sample}\\ 0 & \textit{otherwise.}\end{cases}$$

$$\beta_i=\begin{cases}1 & \text{if } \textit{item i responds given it was sampled}\\ 0 & \textit{otherwise.}\end{cases}$$

The random variables will be assumed independent. In the first part of this note conditioning will be done on the realized values of α_i, thus $\bar{Y}_{HT}$ becomes a fixed quantity.

$\hat{\bar{Y}}_{HT}$ can be written as

$$\hat{\bar{Y}}_{HT}=\frac{1}{N}\sum_{i\varepsilon s}\beta_i\frac{y_i}{\pi_i\,\hat{\rho}_i}$$

By the definition of ρ_i, $E(\beta_i)=\rho_i$. Suppose conditioning is done on the realized $\hat{\rho}_i$'s as well as α_i. Let α be the vector of $\{\alpha_i\}$ and $\hat{\rho}$ the vector of $\{\hat{\rho}_i\}$. Then

$$E(\hat{\bar{Y}}_{HT}\mid\alpha,\hat{\rho})=\frac{1}{N}\sum_{i\varepsilon s}E(\beta_i\mid\alpha,\hat{\rho})\frac{y_i}{\pi_i\,\hat{\rho}_i}.$$

$$=\frac{1}{N}\sum_{i\varepsilon s}\frac{y_i}{\pi_i}\frac{\rho_i}{\hat{\rho}_i}$$

If $\hat{\bar{Y}}_{HT}$ is viewed as an estimator of $\bar{Y}_{HT}$, then

$$Bias\,(\hat{Y}_{HT} \mid \alpha, \hat{\rho}) = E\,(\hat{Y}_{HT} - Y_{HT} \mid \alpha, \hat{\rho}) = \frac{1}{N} \sum_{i \varepsilon s} \frac{y_i}{\pi_i} \left(\frac{\rho_i}{\hat{\rho}_i} - 1 \right).$$

An estimator of this quantity which is unbiased given α and $\hat{\rho}$ is

$$\hat{Bias}\,(\hat{Y}_{HT} \mid \alpha, \hat{\rho}) = \frac{1}{N} \sum_{i \varepsilon rs} \frac{y_i}{\pi_i \hat{\rho}_i} - \frac{1}{N} \sum_{i \varepsilon s} \frac{y_i}{\pi_i}$$

This can be used for any y_i's known for the full sample. This will allow for evaluation of the procedure to estimate $\hat{\rho}_i$ by using frame data.

The variance of $\hat{Bias}\,(\hat{Y}_{HT} \mid \alpha, \hat{\rho})$ is easy to get-

$$Var\,(\,\hat{Bias}\,(\,\hat{Y}_{HT} \mid \alpha, \hat{\rho}\,)\,) = Var\left(\frac{1}{N} \sum_{i \varepsilon rs} \frac{y_i}{\pi_i \hat{\rho}_i} \right)$$

since the second term is fixed,

$$= Var\left(\frac{1}{N} \sum_{i \varepsilon s} \beta_i \frac{y_i}{\pi_i \hat{\rho}_i} \right)$$

$$= \frac{1}{N} \sum_{i \varepsilon s} Var\,(\beta_i) \frac{y_i^2}{\pi_i^2 \hat{\rho}_i^2}$$

assuming the β_i's are independent.

$$= \frac{1}{N} \sum_{i \varepsilon s} \left(\rho_i - \rho_i^2 \right) \frac{y_i^2}{\pi_i^2 \hat{\rho}_i^2}$$

ρ_i is unknown, thus the estimator of $Var\;(\hat{Bias})$ that will be used is

$$\hat{Var}\,(\hat{Bias}\,(\hat{Y}_{HT} \mid \alpha, \hat{\rho})) = \frac{1}{N} \sum_{i \varepsilon s} \left(\hat{\rho}_i - \hat{\rho}_i^2 \right) \frac{y_i^2}{\pi_i^2 \hat{\rho}_i^2}$$

The bias of the variance estimator is

$$E(\hat{Var} - Var) = \frac{1}{N} \sum_{i \in s} [(\hat{\rho}_i - \hat{\rho}_i^2) - (\rho_i - \rho_i^2)] \frac{y_i^2}{\pi_i^2 \hat{\rho}_i^2}$$

Since we've conditioned on the realized values of $\hat{\rho}_i$ the difference between $\hat{\rho}_i$ and ρ_i is a fixed quantity. One can check if the Bias is significantly different from 0.

The next task is calculate and estimate the squared bias of $\hat{Y}_{HT}$.

$$Bias^2(\hat{Y}_{HT} \mid \alpha, \hat{\rho}) = E^2(\hat{Y}_{HT} - Y_{HT} \mid \alpha, \hat{\rho})$$

$$= E((\hat{Y}_{HT} - Y_{HT})^2 \mid \alpha, \hat{\rho})$$

$$-Var(\hat{Y}_{HT} - Y_{HT} \mid \alpha, \hat{\rho}).$$

Estimates are needed for both terms. For the first term a random variable is an unbiased estimator of its own expectation, thus $(\hat{Y}_{HT} - Y_{HT})^2$ is an unbiased estimator of $E(\hat{Y}_{HT} - Y_{HT})^2$. For the second term

$$Var(\hat{Y}_{HT} - Y_{HT} \mid \alpha, \hat{\rho}) = Var(\hat{Y}_{HT} \mid \alpha, \hat{\rho})$$

since Y_{HT} is fixed given $\alpha, \hat{\rho}$. From page 3,

$$Var(\hat{Y}_{HT} \mid \alpha, \hat{\rho}) = \frac{1}{N} \sum_{i \in s} (\rho_i - \rho_i^2) \frac{y_i^2}{\pi_i^2 \hat{\rho}_i^2}$$

and an estimator of this (not unbiased) is

$$\hat{Var}(\hat{Y}_{HT} \mid \alpha, \hat{\rho}) = \frac{1}{N} \sum_{i \in s} (\hat{\rho}_i - \hat{\rho}_i^2) \frac{y_i^2}{\pi_i^2 \hat{\rho}_i^2}$$

The overall estimator of $Bias^2$ is thus

$$\hat{Bias}^2\left(\hat{Y}_{HT} \mid \alpha, \hat{\rho}\right) = \left(\hat{Y}_{HT} - Y\right)^2 - \frac{1}{N}\sum_{i\varepsilon s}\left(\hat{\rho}_i - \hat{\rho}_i^{\,2}\right)\frac{y_i^{\,2}}{\pi_i^{\,2}\,\hat{\rho}_i^{\,2}}$$

Its bias is equal to the bias of $\hat{Var}$.

The next section looks at integration over all possible samples, i.e., over α. On page 2 $E\left(\hat{Y}_{HT} \mid \alpha, \hat{\rho}\right)$ was calculated, with its corresponding bias -

$$E\left(\hat{Y}_{HT} \mid \alpha, \hat{\rho}\right) = \frac{1}{N}\sum_{i\varepsilon s}\frac{y_i}{\pi_i}\frac{\pi_i}{\hat{\pi}_i}$$

$$Bias\left(\hat{Y}_{HT} \mid \alpha, \hat{\rho}\right) = \frac{1}{N}\sum_{i\varepsilon s}\frac{y_i}{\pi_i}\left(\frac{\rho_i}{\hat{\rho}_i} - 1\right)$$

Integrating over α and assuming $\hat{\rho}_i$ is independent of α_i –

$$E\left(\hat{Y}_{HT} \mid \hat{\rho}\right) = \frac{1}{N}\sum_{i=1}^{N} y_i\,\frac{\rho_i}{\hat{\rho}_i}$$

$$Bias\left(\hat{Y}_{HT} \mid \hat{\rho}\right) = \frac{1}{N}\sum_{i=i}^{N} y_i\left(\frac{\rho_i}{\hat{\rho}_i} - 1\right).$$

This is caused by misestimation of $\hat{\rho}_i$ as an estimator of ρ_i. $\hat{Bias}\left(\hat{Y}_{HT} \mid \alpha, \hat{\rho}\right)$ is also an unbiased estimator of $\hat{Bias}\left(\hat{Y}_{HT} \mid \hat{\rho}\right)$, thus

$$\hat{Bias}\left(\hat{Y}_{HT} \mid \hat{\rho}\right) = \frac{1}{N}\sum_{i\varepsilon rs}\frac{y_i}{\pi_i\,\hat{\rho}_i} - \frac{1}{N}\sum_{i\varepsilon s}\frac{y_i}{\pi_i}$$

The final task is to calculate the variance of $\hat{Y}_{HT}$ integrating over all samples -

$$Var\left(\hat{Y}_{HT} \mid \hat{\rho}\right) = Var\left(\frac{1}{N} \sum_{i \varepsilon rs} \frac{y_i}{\pi_i \hat{\rho}_i} \mid \hat{\rho}\right)$$

$$= Var\left(\frac{1}{N} \sum_{i=1}^{N} \alpha_i \beta_i \frac{y_i}{\pi_i \hat{\rho}_i} \mid \hat{\rho}\right)$$

Assuming $\alpha_i \beta_i$ is independent of $\alpha_i \beta_i$,

$$= \frac{1}{N^2} \sum_{i=1}^{N} Var\,(\alpha_i \beta_i) \frac{y_i^{\,2}}{\pi_i^{\,2} \hat{\rho}_i^{\,2}}$$

$$Var\,(\alpha_i \beta_i) = \mathrm{E}\,(\alpha_i^{\,2} \beta_i^{\,2}) - [\mathrm{E}\,(\alpha_i \beta_i)]^2$$

Assuming α_i and β_i are independent and noting that $\alpha_i^{\,2} = \alpha_i$, $\beta_i^{\,2} = \beta_i$ since they are indicator random variables,

$$Var\,(\alpha_i \beta_i) = \pi_i \rho_i - \pi_i^{\,2} \rho_i^{\,2}$$

Thus

$$Var\left(\hat{Y}_{HT} \mid \hat{\rho}\right) = \frac{1}{N^2} \sum_{i=1}^{N} (\pi_i \rho_i - \pi_i^{\,2} \rho_i^{\,2}) \frac{y_i^{\,2}}{\pi_i^{\,2} \hat{\rho}_i^{\,2}}$$

An unbiased estimator of this is

$$\hat{Var}\,(\hat{Y}_{HT} \mid \hat{\rho}) = \frac{1}{N^2} \sum_{i \varepsilon s} \frac{\pi_i \rho_i - \pi_i^{\,2} \rho_i^{\,2}}{\pi_i} \frac{y_i^{\,2}}{\pi_i^{\,2} \hat{\rho}_i^{\,2}}$$

This ass''mes ρ_i is known. Now replace ρ_i by $\hat{\rho}_i$ to achieve an estimator of $Var\,(\hat{Y}_{HT} \mid \hat{\rho})$ which is calcuulable -

$$\hat{Var}(\hat{Y}_{HT} \mid \hat{\rho}) = \frac{1}{N^2} \sum_{I \varepsilon s} \frac{\pi_i \hat{\rho}_i - \pi_i^2 \hat{\rho}_i^2}{\pi_i} \frac{y_i^2}{\pi_i^2 \hat{\rho}_i^2}$$

This will be biased, with bias

$$\frac{1}{N^2} \sum_{i \varepsilon s} \left(\pi_i \hat{\rho}_i - \pi_i^2 \hat{\rho}_i^2\right) - \left(\pi_i \rho_i - \pi_i^2 \rho_i^2\right) \left[\frac{y_i^2}{\pi_i^3 \hat{\rho}_i^2}\right]$$

All of the above computations can be made with the simple non-response adjustment, in which $\hat{\rho}_i$ will be the overall response rate, or overall response rate in the strata, or with the model adjustment, with $\hat{\rho}_i$'s obtained from logistic regression.

These computations can now be used to evaluate the various estimators under non-response, using y_i 's known for the full sample.

Section 2

To estimate the probabilities of responding,logistic regression was carried through with response as the dependent random variable-

$$y_i = \begin{cases} 1 \text{ if } \textit{school responded survey} \\ 0 \; \textit{otherwise} \end{cases}$$

The following model is assumed for y_i

$$y_i \sim Bern(\pi_i) \text{ , } (i.e., y_i \sim Bern(\pi_i, 1))$$

$$\text{with} \qquad \ln \frac{\pi}{(1-\pi_i)} = X_i{}' \beta$$

where x_i is a vector of covariates and β is a parameter vector. Maximum Likelihood is used to estimate β. Once $\hat{\beta}$ is estimated using Maximum Likelihood the predicted probability of responding for school i is

$$\hat{\pi}_i = \frac{\exp(x_i \, \beta hat)}{L + \exp(x_i \, \beta hat)}$$

(See McCullagh & Nelder (1983) for a description of the theory underlying these models and calculations)

The main question of concern is how to select the model-i.e. how to choose the covariate structure. We are limited to using covariates that are known for all schools sampled, both responding and non-responding. If a covariate is known only for responding schools then the parameter associated with it cannot be estimated. Thus covariates collected from the schools after response are not useful.

There is available a set of covariates known for all schools in the population, such as whether they are public or private, urban or rural, and where they are located in the country, among others. Also a follow-up study was taken of schools that did not respond. The questions asked in the follow-up study were selected based on the supposition that they were correlated to the response behavior of the school. The response rate for the follow-up study was very high, thus this study provides a new set of covariates known for the full sample of both responding and non-responding schools. Along with the frame covariates we have a rich set of covariates that can be used to predict response.

Only Pool 1 schools were used in the final model, since only these schools were included in the follow-up study. Certain augmentation schools were also excluded.

Using all of the covariates would be unwieldy, thus a subset must be chosen. A clean way to do this would be to use stepwise techniques. Stepwise logistic regression is possible using SAS, but it is very expensive. Thus for the stepwise search normal regression was used rather than logistic regression, i.e. a simple regression model

$$y_i = x_i \, \beta + e_i$$

was used with the stepwise techniques. Generally the p-values of the coefficients of a given model tend to be very close when the same set of covariates was fit using both models, thus the normal model approximates the logistic model adequately with this data when a covariate selection is being done. For example, with the final covariate set the p-values associated with the parameter estimates for the two models were

Logistic Model		Normal Model	
Q4P1	.0192	Q4P1	.0211
Q40P1	.0092	Q40P1	.0100
Q40P2	.0370	Q40P2	.0363
Q41P15	.0281	Q41P15	.0221
Q24P4	.0001	Q24P4	.0001
Q37	.0001	Q37	.0001
PROB	.3615	PROB	.3717
SIZE	.5929	SIZE	.5638

The stepwise regression technique used was a backwards elimination with entry and exit p-values of .25 (see SAS manual-Statistics).

The very liberal entry and exit cut-offs allow for a very large model-in this case bias results if a covariate is left out of the model mistakenly, thus the conservative approach would not be alpha-level acceptance of the null hypothesis of a zero coefficient. Some covariates were forced into the model, such as urbanicity, and size of the school.

The stepwise procedure used was PROC STEPWISE on SAS. The covariates were not all added at once, but in a small number of large groups. Interactions were checked between covariates that remained in the final model. For all covariates the marginal frequency tables of the covariates against response were checked to make sure a significant coefficient was not due to a very small amount of data-for example a situation such as

Covariate A	Level 1	800	200	80% response
	Level 2	5	5	50% response

might generate a significant coefficient for covariate A in the model because of the great differential in response rates, but the difference is due to the limited amount of data at level 2.

Any significant coefficients due to a small amount of data were thus rejected. The final model selected was then fit using logistic regression with PROC LOGIST on SAS-The model fit is as folllows -

VARIABLE	BETA	STD. ERROR	CHI-SQUARE	P
INTERCEPT	-1.73779159	0. 82635595	4.42	0.0355
Q4P1	-0.56678020	0.24195014	5.49	0.0192
Q40P1	0.46913561	0.17998395	6.79	0.0091
Q40P2	0.38450315	0.18437734	4.35	0.0370
Q41P15	-0.75456755	0.34371827	4.82	0.0281
Q24P4	1.57827111	0.40579189	15.13	0.0001
Q37	-0.34687118	0.08267074	17.60	0.0000
PROB	0.87740465	0.96149825	0.83	0.3615
SIZE	0.00040207	0.00075195	0.29	0.5929

The covariates are defined as follows:

Q4P1 =	0	If school is private
	1	If school is public
Q40P1 =	0	If school is non-urban
	1	If school is urban
Q40P2 =	0	If school is non-rural
	1	If school is rural
Q41P15 =	0	If otherwise
	1	If school is from West South Atlantic region
Q24P4 =	0	If otherwise
	1	If pupil was assigned to this school based on academic criteria
Q37 =	1	Standardized test results always provided to family
	2	Stnadardized test results usually provided to family
	3	Standardized test results sometimes provided to family
	4	Standardized test results seldom provided to family
	5	Standardized test results never provided to family

PROB = the probability the school has of being selected into the sample.
SIZE = the number of students in the eighth grade class.

As a means to understan ` `he model, the following pages give the marginal response rates given the level of the covariates in the model. For type of school, the marginal frequency table was:

		Public	Private	
Responded?	0	49 23.79%	249 32.63%	298 30.75%
	1	157	514	671 69.25%
		206	763	969

The numbers in the cells are the actual frequencies among the 96 schools used to fit the model. The percentages are non-response rates for each type of school. As can be seen, the non-response rate was higher for private schools - 32.69% as against 23.8% for public schools. For the other covariates, the following tables can be interpreted the same way.

For urbanicity -

		Urban	Suburban	Rural	
Responded?	0	71 23.8%	144 35.0%	83 27.8%	298 30.8%
	1	214	267	190	671
		285	411	273	969

Non-response in suburban schools was higher than either urban or rural schools. For geography:

		West South Atlantic	Other Region	
Responded?	0	19 50.0%	279 30.0%	298 30.8%
	1	19	652	671
		38	931	969

For Q24P4 -

		Uses Achievement Test for Entrance	Other	
Responded?	0	18 62.1%	280 29.8%	298 30.8%
	1	11	660	671
		29	940	969

The last qualitative covariate is Q37 -

		1	2	3	4	5	
Responded?	0	222 30.0%	20 15.6%	22 41.5%	30 75.0%	4 57.1%	298 30.8%
	1	519	108	31	10	3	671
		741	128	53	40	7	969

The model was fit only using POOL1 data. For other schools with covariate values for all the covariates in the model. Predictions were made based on the model. There were some schools with covariate values only from the frame - i.e., they had covariate values for Q4P1, Q4OP1, Q4OP2, Q41P15, PROB, and SIZE. For these schools, a logistic regression model was fit with these covariates only, and predictions of non-response propensity were generated for these schools based on the smaller model.

The predictions of the response propensity for school in the sample based on the mode range from a minimum of .330 to a maximum of .904. The percentiles were as follows -

MINIMUM	33.0%
1st Percentile	34.5%
5th Percentile	47.6%
10th Percentile	55.5%
25th Percentile	65.9%
MEDIAN	71.8%
75th Percentile	76.1%
90th Percentile	81.4%
95th Percentile	83.0%
99th Percentile	85.7%
MAXIMUM	90.5%

These predicted probabilities of response will be the $\hat{\rho}_i$ ' s used in the adjusted estimator in Section 1.

Finally, based on preliminary work of checking the adjustment with the logistic regression $\hat{\rho}_i$'s a few general comments can be made. For any response variable, highly correlated to a covariate used in the logistic model of adjustment should reduce non-response bias quite well. For ar response variable highly correlated to the covariate frame the frame or foilow-up study, l out the model and therefore is little non-response bias to adjust for, and these response variables will have little non-response bias in the adjusted estimates. Any response variables that are not correlated to any frame or follow-up covariates are not adjusted - if there is a relationship between these response variables and whether the school responds, there will be non-response bias unadjusted for by the adjusted estimators.

These ideas ought to be checked using frame variables known for all units.

III. Item Nonresponse Analysis - Steven Ingels, Kenneth Rasinski

Analysis of survey error is important for understanding potential bias in making inferences from an obtained sample to a population. Sampling and nonsampling errors are the key constituents of total survey error. in the NELS:88 Sample Design Report sampling error analyses for the Base Year document design effects and standard errors for key variables. Non-sampling error--the bias associated with unit and item nonresponse--must also be described and quantified.

In a two-stage sample such as the NELS:88 Base Year, one type of nonsampling error, unit nonresponse, can occur at either stage. Unit nonresponse can occur at the first selection stage when a school declines to participate, or at the second stage when an individual respondent within a participating school does not participate. This report documents the magnitude and effect of nonresponse at the first, or school, level of sampling, and makes inferences about the effect of the second level. Item nonresponse occurs when a respondent fails to complete certain items on the survey instrument. Of course, as we move from the data in its pristine form, to intermediate and final public use versions of a datatape, item nonresponse can either be increased (see Ingels, 1987, for a prominent example) or diminished (see the discussion of NELS:88 machine editing and composite variables below) as an artifact of the file construction process.

While bias associated with unit nonresponse at both the school and the individual level has been controlled by making adjustments to case weights, item nonresponse has generally not been compensated for in the NELS:88 data set. There are two partial exceptions to this generalization. The first partial exception is machine editing, through which, occasionally, certain nonresponse problems are rectified by imposing interitem consistency, particularly by forcing logical agreement between filter and dependent questions. Thus, for example, the missing response to a filter question can often be inferred if the dependent question has been answered. In addition to replacing a missing datum with the inferred "correct" datum, machine editing also can reduce other sorts of nonresponse, such as certain instances in which a "mark one response only" question has illegitimate multiple response. If the response categories invoke a hierarchy, a decision rule can be formulated such that the highest level marked should be taken as the sole valid response to the question. (An example of such a hierarchical question would be the various NELS:88 items that ask "What is the highest level of education that you/your spouse/your mother/your father have completed.")

The second partial exception is that some key student classification variables have been constructed in part from additional sources of information when student data are missing. Thus, data from school records (for example, student sex or race/ethnicity as given on the sampling roster) or from the parent or teacher questionnaire (for example, limited English proficiency status) have been used to replace missing student data. However, apart from these special cases, missing values have not been imputed in the NELS:88 data. Since item nonresponse is an important potential and uncorrected source of data bias, it is necessary to measure its impact so that analysts can properly take potential response biases into account.

There are two main purposes to this analysis. One purpose is to quantify nonresponse bias for key variables on the student questionnaire and tests. A second purpose is to describe nonresponse patterns, both in terms of characteristics of items and in terms of characteristics of respondents. The result is an analysis of the maximum and probable bias due to item nonresponse, identifying items for which data analysts should exercise caution in generalizing from NELS:88 data.

The present item nonresponse analysis employed the machine-edited data, not the original raw data. Our aim was to convey to users the possible item nonresponse bias in the final public use version of the data. Had our purpose been, as for example it was in the field test, the revision of items in the questionnaire so that future nonresponse problems could be minimized, we would have employed the original--unedited--datafiles.

The analysis proceeded in three stages. In the first stage, average nonresponse rates were calculated for each item. In the second stage, nonresponse was evaluated as a function of item characteristics: (1) position in the questionnaire, (2) topic, and (3) whether the item was contingent on a filter. Items with relatively high nonresponse rates were selected for further analysis in stage two. In the third stage, nonresponse rates for selected high nonresponse items and for test scores were modeled as a function of respondent characteristics. While analysis of variance was used to examine differences for statistical significance, it should be noted that it is unlikely that the nonresponse measures used as dependent variables are normally distributed, and often the sample sizes of the groups comprising levels of various factors are very different.

III-A. Population and data file definitions.

DEFINITION 1: "ITEM."

For purposes of this analysis, "item" refers to each data element or variable. For a question composed of multiple subparts, each subpart eliciting a distinct response is counted as an item for item nonresponse purposes. (Thus, a single question that poses three subquestions is treated as three variables).

DEFINITION 2: "RESPONSE RATE."

NCES standards (NCES, 1987) stipulate that item response rates (Ri) "are to be calculated as the ratio of the number of respondents for which an in-scope response was obtained (i.e., the response conformed to acceptable categories or ranges), divided by the number of completed interviews for which the question (or questions if a composite variable) was intended to be asked.":

$$Ri = \frac{\text{weighted \# of respondents with in-scope responses}}{\text{weighted \# of completed interviews for which question was intended to be asked.}}$$

In-scope responses were considered to be valid answers (including a "don't know" response when this was a legitimate response option.) Out-of-scope responses were multiple responses to items requiring only a single response, refusals, and missing responses.

DEFINITION 3: "STUDENT POPULATION."

A. Item nonresponse analysis population, student questionnaire. All students who completed the questionnaire, regardless of whether they completed the test.

B. Student nonresponse analysis population, student test. Test + questionnaire cases are the only cases included; test-only cases have been intentionally excluded; and, of course, "no test" cases, and nonrespondents, necessarily fall outside the analysis.

DEFINITION 4: "STUDENT DATA."

Student questionnaire data file. The public use datafile with machine-edited, weighted data was used as the basis for the analysis. Nonresponse rates of composite and other constructed variables were not examined in this analysis.

Student test datafile. The weighted datafile for the four tests in the NELS:88 cognitive test battery.

III-B. Quantification of Item Nonresponse.

Item-level nonresponse. Weighted nonresponse rates equal to the proportion of eighth graders who failed to answer a particular item (that is, 1-Ri) were calculated for each item in the student questionnaire. The average item nonresponse is 4.7 percent (standard deviation, 3.5 percent). Items deviate markedly from this average. For some items nonresponse is zero. For other items the nonresponse rate is as high as 21.6 percent.

Table III-1 shows statistics for the item nonresponse rates overall and for items grouped into categories depending upon their position in the questionnaire, the topic they addressed, and whether they were part of a skip or filter pattern. When items were grouped into thirds based on their serial position in the questionnaire, mean nonresponse rates differed significantly across thirds ($F(2,273)=79.29$, $p<.01$). A slightly higher nonresponse rate is found for items near the beginning of the questionnaire, and a substantially higher nonresponse rate is found for items near the end of the questionnaire.

Table III-1. Statistics on Proportion Nonresponding by Various Item Characteristics[1]

Domain	Average	Standard Deviation	Minimum	Maximum	Number of Items[2]
Overall	.047	.035	.000	.216	276
Position[3]					
First Third	.037	.037	.000	.216	72
Second Third	.028	.018	.000	.094	107
Last Third	.075	.028	.026	.167	97
Last Third without last two sets	.064	.028	.026	.167	66
Topic					
Student Background	.030	.038	.008	.135	10
Language Use	.050	.033	.002	.143	26
Family	.034	.037	.000	.216	50
Self-Concept	.016	.004	.010	.022	13
Future Plans	.025	.014	.000	.058	37
Jobs and Chores	.009	.013	.000	.018	2
School Life	.029	.007	.017	.048	38
School Work	.063	.028	.026	.167	69
Student Activities	.098	.007	.083	.115	31
Filtered					
Yes	.058	.045	.008	.216	32
No	.045	.033	.000	.167	244

1 All values are based on weighted data.

2 The number of items used in this analysis is the total number of items in the student questionnaire minus those items that were part of a "mark all that apply" sequence. These "mark all that apply" items were excluded because it was impossible to distinguish a response indicating the item did not apply from a nonresponse.

3 Unequal numbers of items in each of the thirds result because items were divided into thirds before the "mark all that apply" items were excluded. This practice served to preserve the equal cutting of the questionnaire into thirds regardless of whether each item in each of the thirds was used in the analysis.

The last two sets of items require students to indicate their participation in a number of activities. It is possible that fatigue effects, naturally occurring at the end of a long questionnaire, may be exacerbated by these somewhat tedious questions, accounting for most of the differences in the last third of the questionnaire. When nonresponse rates for the last two sets of items are compared with nonresponse rates for all the preceding items, average nonresponse for the last two sets is significantly higher (31 items, average proportion nonresponding=.098, versus all preceding items, average proportion nonresponding=.040, $F(1,274)=160.96$, $p<.01$). A reanalysis of nonresponse rates across thirds of the questionnaire was conducted after removing the last two items. Differences among thirds were less, but still significant ($F(2,242)=35.55$, $p<.01$) suggesting that the last two sets of items account for some but not all of the higher nonresponse at the end of the questionnaire. Post hoc Neuman-Keuls analysis indicates that the nonresponse rates for the first two thirds are not significantly different from each other, while the nonresponse rate for the last third (either with or without the final two sets of items included) differs significantly from the rates for the other two thirds. Nonresponse rates for the various configurations of serial position are presented in Table III-1.

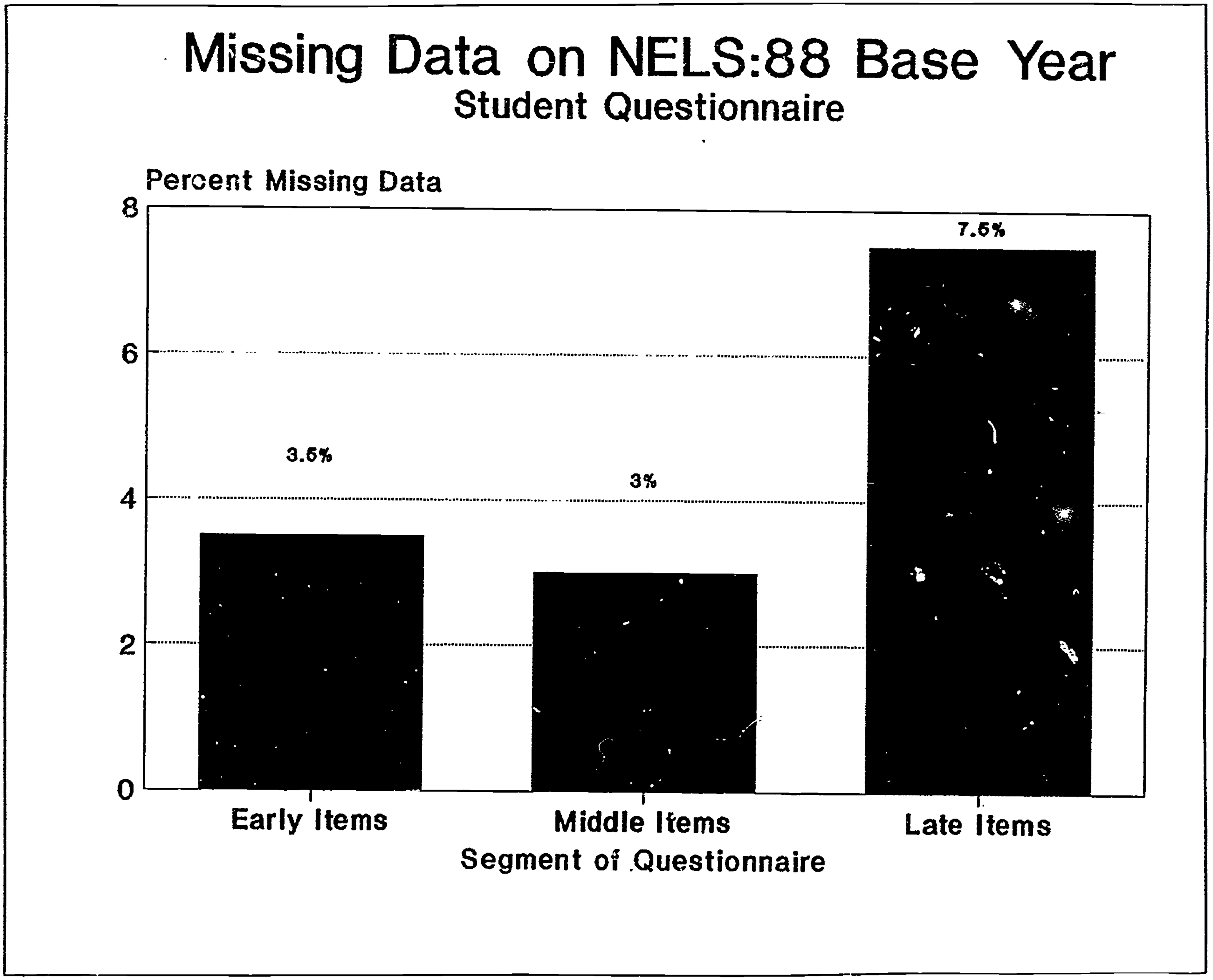
While nonresponse rates indeed rise in the last third of the NELS:88 questionnaire, they are nonetheless modest compared to those of the base year of High School and Beyond--22 percent in HS&B, 7.5 percent in NELS:88--see Figures III-1 and III-2 for comparisons of early, middle and late items in the respective base years of the two surveys.

The NELS:88 Base Year student questionnaire was constructed such that questions in each of the nine sections formed topical blocks. Table III-1 also shows the average nonresponse rates by topic. The difference by topic is significant ($F(8,267) = 29.83$, $p < .01$), however, the substantially discrepant numbers of items in each of these categories, ranging from 2 to 69 items, suggests a cautious interpretation. Post hoc Neuman-Keuls tests indicate that nonresponse rates for questions on student participation in activities are significantly higher than nonresponse rates for other topics. The section comprising questions about language use differed significantly in nonresponse from the sections on self-esteem/locus of control and jobs and chores. The remaining sections did not differ significantly from one another.

Item nonresponse was also analyzed as a function of whether the item was part of a filter-dependent question. Thirty-two items were of this type, and nonresponse for these items was compared to the two hundred and forty four items that were not in a dependent relationship with a filter item. As Table III-1 shows, there is a slightly higher nonresponse rate for items that were filtered than for those that were not. This difference was significant ($F(1,275) =4.00$, $p<.05$).

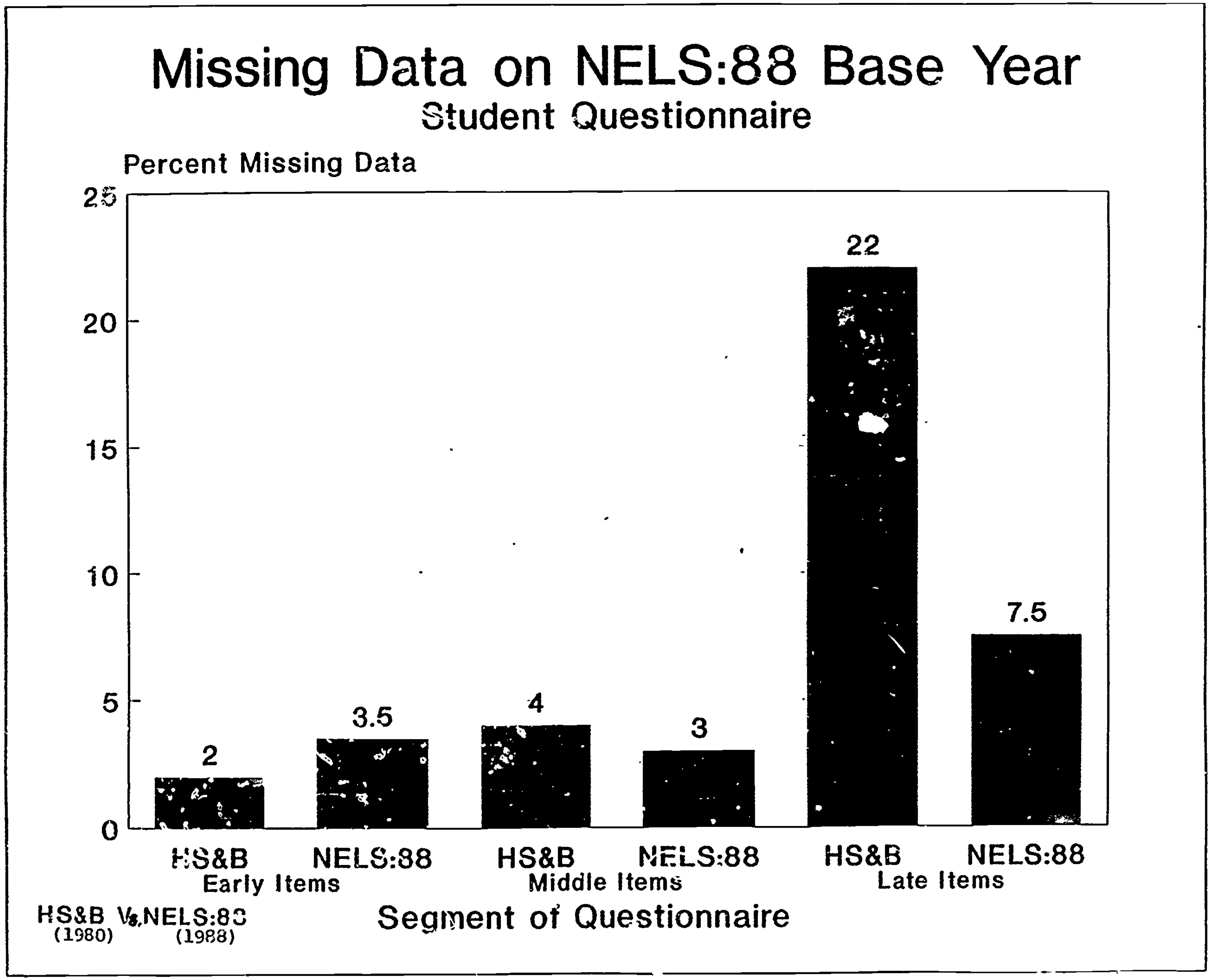
Critical Items. A number of items in the student questionnaire were dubbed "critical items" because of their special interest to analysts, their policy relevance, or their usefulness in locating the student for subsequent follow-up studies. These items were edited by the NORC field personnel who administered the survey. If the response to one or more of the critical items was missing, undecipherable, or had multiple categories marked when only one response was required, the NORC field staff member privately pointed out the problem to the student. If, after prompting, the student indicated that he or she had chosen not to answer the question, the NORC staff member marked a "no retrieval" response for the item. ("No retrieval" was indicated by filling in an oval positioned to the left of each critical item). The "no retrieval" responses were used later during the machine editing process to assign a "refused" response to the critical items. Most editing and retrieval for the student questionnaire was conducted in the way just described. In a very small

Figure III-1



34

Figure III-2



Missing Data on NELS:88 Base Year
Student Questionnaire
Percent Missing Data
25
20
15
10
5
0
2
3.5
4
3
22
7.5
HS&B
NELS:88
HS&B
NELS:88
HS&B
NELS:88
Early Items
Middle Items
Late Items
HS&B Vs.NELS:88
(1980)
(1988)
Segment of Questionnaire

number of instances (fewer than 300 cases), additional retrieval of missing responses to critical items had to be conducted after the questionnaire reached NORC.

The average item nonresponse rate for each of the critical items is shown in Table III-2. The items in this table represent the majority but not the total set of critical items. Critical items that were part of the locator information were excluded from this analysis. Nonresponse was quite high for one critical item--the question asking about Asian/Pacific Islander ethnic subgroup shows a nonresponse rate of just over one percent until adjustment is made for those who legitimately skipped the question. Once those who were routed out are removed from the denominator, the adjusted nonresponse rate climbs to 21.6 percent. Otherwise, no critical item is as high as 10 percent, with most being considerably less (24 of the 42 critical data elements are at 2.6 percent nonresponse or less, 15 are at two percent or lower, while five are at one percent nonresponse or lower).

Table III-2. Average Proportion Nonresponding to Critical Items

Item Number Nonresponse	Topic	Average
BYS2A	Is R's mother living?	0.017
BYS4A	Current job status of R's mother	0.014
BYS5A	Is R's father living?	0.019
BYS7A	Current job status of R's father	0.040
BYS8	People in R's household	0.012
BYS12	Respondent's sex	0.008
BYS21	Is a language other than English spoken in R's home?	0.002
BYS22	Language usually spoken at R's home	0.046
BYS31A	Respondent's race	0.010
BYS31B	Asian or Pacific Islander subcategory	0.216
BYS31C	Hispanic subcategory	0.087
BYS31D	Hispanic race	0.079
BYS34A	Father's level of education	0.000
BYS34B	Mother's level of education	0.000
BYS51AA	Talked to counselor about high schools	0.014
BYS51AB	Talked to teacher about high schools	0.019
BYS51AC	Talked to other adult about high schools	0.019
BYS51BA	Talked to counselor about jobs/careers	0.020
BYS51BB	Talked to teacher about jobs/careers	0.025
BYS51BC	Talked to other adult about jobs/careers	0.018
BYS51CA	Talked to counselor to improve academic work	0.025
BYS51CB	Talked to teacher to improve academic work	0.018
BYS51CC	Talked to other adult to improve academic work	0.026
BYS51DA	Talked to counselor about course selection	0.024
BYS51DB	Talked to teacher about course selection	0.029
BYS51DC	Talked to other adult about course selection	0.030
BYS51EA	Talked to counselor about class-work	0.038
BYS51EB	Talked to teacher about class-work	0.029
BYS51EC	Talked to other adult about class-work	0.035
BYS51FA	Talked to counselor because of discipline problems	0.042
BYS51FB	Talked to teacher because of discipline problems	0.045
BYS51FC	Talked to other adult because of discipline problems	0.043
BYS51GA	Talked to counselor about alcohol or drug abuse	0.023
BYS51GB	Talked to teacher about alcohol or drug abuse	0.027
BYS51GC	Talked to other adult about alcohol or drug abuse	0.026
BYS51HA	Talked to counselor about personal problems	0.025
BYS51HB	Talked to teacher about personal problems	0.033
BYS51HC	Talked to other adult about personal problems	0.024
BYS81A	Grades in English from 6th grade up till now	0.026
BYS81B	Grades in math from 6th grade up till now	0.028
BYS81C	Grades in science from 6th grade up till now	0.028
BYS81D	Grades in social studies from 6th grade up till now	0.030

Note: All values are based on weighted data.

III-C Individual differences in nonresponse: demographic characteristics of individuals with a greater item nonresponse propensity

Individual differences in nonresponse. Nine questions with the highest nonresponse rates were selected for analysis to determine the relationship between nonresponse and student characteristics. These questions and their nonresponse rates are listed in Table III-3. Table III-4 shows the proportion nonresponding to the nine items with the highest nonresponse rates by selected student characteristics. A composite nonresponse variable was created by counting the number of items for which a nonresponse was given across items 24, 29, 67A, 67C, 67AA, 67AC, 67AD, and 83J from Table III-3 (the high nonresponse items available for the full sample of students) for each student. This composite, which could range from zero to six, was used as a dependent variable in an analysis of variance, with the student's sex, racial/ethnic background, socioeconomic status, and test composite quartile as independent variables. The analysis of variance examined nonresponse as a function of main effects only, ignoring interactions among the independent variables.

Results of this analysis suggest that boys were significantly more likely to be nonrespondents on these items than girls ($F(1,23459)=143.17$, $p<.01$). The analysis also indicates that there are significantly different nonresponse rates across the five racial/ethnic groups ($F(4,23459)=50.68$, $p<0.0001$). Post hoc Neuman-Keuls tests indicate that blacks were most likely to be nonrespondents, averaging nonresponse to 1.509 items across the six item scale. Hispanics were next most likely, averaging 1.127 nonresponding items. Asians and American Indians were third, averaging .9481 and .9454 items respectively, but not differing between them. Finally, whites had the least tendency toward nonresponse, averaging .7439 items. A single degree-of-freedom linear contrast of nonresponse across the four test quartiles was significant, indicating that students with lower test scores were more likely to be nonrespondents than those

Table III-3. Nine Items with the Highest Nonresponse Rates

		Proportion Nonresponding	Eligible Respondents
BYSQ16.	[IN REFERENCE TO A SECOND NOMINATED HIGH SCHOOL] Is this a public school. a private religious school, or a private nonreligious school?	0.137	6,687
BYSQ24.	What language, other than English, do you currently use most often?	0.146	5,655
BYSQ29.	Were you ever enrolled in an English language/language assistance program, that is, a program for students whose native language is not English?	0.120	5,655
BYSQ67A.	Which of the following math classes do you attend at least once a week this school year?--Remedial math	0.168	24,599
BYSQ67C.	Which of the following math classes do you attend at least once a week this school year?--Algebra (or other advanced math)	0.135	24,599
BYSQ67AA.	Which of the following science classes do you attend at least once a week this school year?--A science course in which you have a laboratory	0.137	24,599
BYSQ67AC.	Which of the following science classes do you attend at least once a week this school year?--Biology (life science)	0.144	24,599
BYSQ67AD.	Which of the following science classes do you attend at least once a week this school year?--Earth Science	0.114	24,599
BYSQ83J.	Have you or will you have participated in any of the following outside-school activities this year, either as a member, or as an officer (for example, vice-president, coordinator, team captain)?--OTHER	0.117	24,599

Note: Proportions were calculated using weighted data.

Table III-4. Proportion Nonresponding to Nine Items with Highest Nonresponse Rates by Selected Student Characteristics

	Q16	Q24	Q29	Q67A	Q67C	Q67AA	Q67AC	Q67AD	Q83J	Average
Overall	.137	.146	.120	.168	.135	.137	.144	.114	.117	.135
Sex										
Male	.151	.174	.122	.201	.161	.160	.168	.134	.136	.156
Female	.124	.119	.117	.135	.109	.113	.120	.094	.097	.114
Race/ethnicity										
Asian	.147	.144	.059	.183	.138	.144	.154	.129	.129	.136
Black	.116	.301	.221	.272	.246	.241	.244	.196	.216	.228
White	.141	.183	.160	.142	.107	.110	.118	.093	.090	.127
Hispanic	.147	.091	.087	.204	.174	.170	.181	.140	.155	.150
Amerindian	.105	.219	.168	.149	.133	.142	.152	.094	.159	.147
Socioeconomic Status (SES)										
Lowest Quartile	.147	.140	.112	.207	.195	.181	.182	.149	.166	.164
Second Quartile	.135	.135	.106	.160	.141	.134	.141	.117	.123	.133
Third Quartile	.140	.159	.136	.147	.114	.118	.125	.096	.102	.126
Highest Quartile	.122	.157	.132	.157	.091	.113	.124	.094	.076	.118
Cognitive Test Composite										
Lowest Quartile	.172	.194	.149	.237	.238	.213	.221	.184	.198	.201
Second Quartile	.100	.138	.120	.176	.155	.149	.154	.124	.122	.138
Third Quartile	.106	.122	.101	.134	.094	.102	.107	.080	.084	.103
Highest Quartile	.099	.108	.073	.116	.042	.076	.085	.058	.055	.079

Note: Proportions were calculated using weighted data.

with higher test scores (F(1,23459)=476.76, p< 0.0001. A similar test for SES failed to show a significant difference (F(1,23459=0.00, ns). Though the design effect correction was not used in these analyses, it should be noted that the F statistics were large enough that correcting by the average design effect of 2.54 would not have eliminated significant effects.

III-D: Nonresponse on the cognitive test battery

Test Scores. Nonresponse patterns for test scores were analyzed by examining the number of items not attempted for each of the four cognitive tests. These values for the entire student sample and by sex, racial/ethnic, and SES subgroups are shown in Table III-5. Each measure was included in an analysis of variance, with sex, race/ethnicity, and SES as independent variables. As before, interactions were ignored and only main effects were tested. A single degree-of-freedom contrast indicated a significant linear effect by SES for reading F(1,23411)=134.09, p<.01) math (F(1,23395)=51.53, p<.01), history/citizenship (F(1,23295)=28.84, p<.01), and science (F(1,23382)=14.40, p<.01. For all test subjects, lower SES was related to higher nonresponse. Girls showed significantly less nonresponse than boys on math (F(1,23395)=12.13, p<.01) and on reading (F(1,23411)=59.97); however, even though these differences are significant, they are quite small. No significant difference for science or for history/citizenship was found. While the main effect of race/ethnicity was significant for each of the four tests even accounting for the design effect, post hoc Neuman-Keuls tests did not indicate significant differences among the racial groups. Once again, all reported effects are large enough to remain significant even if adjusted down by the average design effect of 2.54.

Another method for assessing test nonresponse is to examine the percent of students who gave an answer to the final item in each test. This has been proposed as an index of test "speededness." Generally, a test is considered to be "unspeeded" if over 80 percent of the test takers attempt the last item. Table III-6 shows that test speededness was not a problem for these broad categories of students. This suggests that the appropriate amount of time was given for completion of each of the four cognitive tests. A detailed account of NELS:88 cognitive test results can be found in the Psychometric Report for the NELS:88 Base Year Test Battery (Rock & Pollack, 1989).

While more respondents attempted the HS&B tests than the NLS-72 tests (Rock et al., 1985, p. 33), the NELS:88 Base Year results show considerable improvement over HS&B in terms of test item response (for HS&B see Table III-7 and Figures III-3, -4 and -5.), as well as overall test reliabilities (particularly for the reading test). In general, for both cognitive tests and student questionnaires, the NELS:88 Base Year achieved a higher level of item response than did the HS&B Base Year, despite the greater heterogeneity in ability levels typical of eighth and earlier grades, which encompass students who are at a modal age temporally prior to the earliest legal school-leaving age. Nevertheless, data users need to be thoroughly aware that missing data was not imputed, and that the problem of missing data is an important potential source of bias.

Table III-5. Average Number of Items Not Attempted on Four Cognitive Tests by Selected Student Characteristics

	Reading	Math	Science	History/ Citizenship	Average
Overall	0.391	0.922	0.437	0.285	0.509
Sex					
Male	0.454	0.978	0.451	0.286	0.542
Female	0.327	0.866	0.422	0.282	0.474
Race/ethnicity					
Asian	0.350	0.812	0.473	0.347	0.496
Black	0.840	1.687	0.751	0.485	0.941
White	0.268	0.718	0.347	0.216	0.387
Hispanic	0.611	1.278	0.577	0.432	0.725
Amerindian	0.578	1.226	0.748	0.461	0.753
Socioeconomic Status (SES)					
Lowest Quartile	0.624	1.228	0.541	0.387	0.695
Second Quartile	0.420	0.984	0.466	0.320	0.548
Third Quartile	0.323	0.833	0.390	0.232	0.445
Highest Quartile	0.201	0.647	0.349	0.198	0.349

Note: Statistics were calculated using weighted data.

Table III-6. Speededness Indices for Test by Racial/Ethnic and Sex Groups (Percent of Sample Who Reached Last Item)

Test	Asian	Hispanic	Black	White	Male	Female
Reading	96.1	92.7	87.9	97.3	94.9	95.9
Math	96.1	93.2	89.7	96.2	95.0	94.9
Science	96.2	95.3	92.6	98.0	96.7	97.0
History/Citizenship	96.2	95.5	94.6	97.9	97.0	97.3

Note: Table excerpted from Rock & Pollack (1989).

TABLE III-7

HS&B Base Year Test Battery

Senior Cohort

Test	Mins Sectn	Total Time	Sectn Items	Total Items	Total N Missing	Total % Missing	Test takers N Missing	Test takers % Missing
1. Vocabulary-1	5	5	15	15	3,324	11.8	153	0.6
2. Vocabulary-2	4	9	12	27	3,363	11.9	192	0.8
3. Reading	15	24	20	47	3,343	11.8	172	0.7
4. Mathematics-1	15	39	25	72	3,473	12.3	302	1.2
5. Mathematics-2	4	43	8	80	3,877	13.7	706	2.8
6. Picture/Number	5	48	15	95	3,895	13.8	724	2.9
7. Mosaic Comp-1	3	51	56	151	4,024	14.2	853	3.4
8. Mosaic Comp-2	3	54	33	184	4,361	15.4	1,190	4.7
9. Visual. in 3D	9	63	16	200	4,370	15.5	1,199	4.8
10. Test Q's	5	68	6	206	5,000	17.7	1,829	7.3

Sophomore Cohort

Test	Mins Sectn	Total Time	Sectn Items	Total Items	Total N Missing	Total % Missing	Test takers N Missing	Test takers % Missing
1. Vocab	7	7	21	21	2,601	8.7	140	0.5
2. Reading	15	22	20	41	2,640	8.8	179	0.6
3. Math-1	16	38	28	69	2,680	8.9	219	0.8
4. Math-2	5	43	10	79	2,950	9.8	489	1.8
5. Science	10	53	20	99	2,910	9.7	449	1.6
6. Writing	10	63	17	11[illegible]	3,163	10.5	702	2.5
7. Civics	5	68	10	126	3,423	11.4	962	3.5

Figure III-3

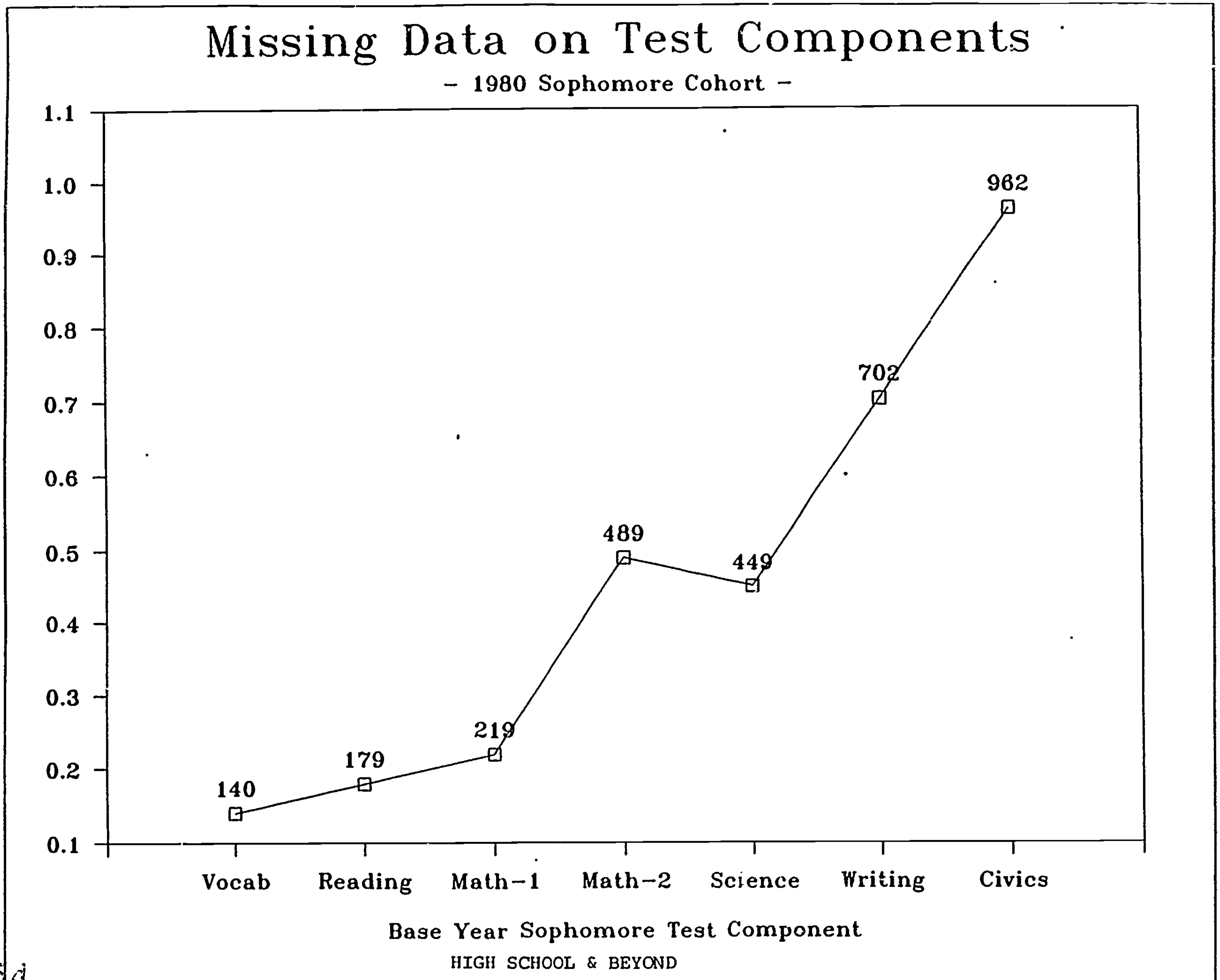

Missing Data on Test Components
- 1980 Sophomore Cohort -
Number Missing Component Score (Thousands)
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
962
702
489
449
219
179
140
Vocab
Reading
Math-1
Math-2
Science
Writing
Civics
Base Year Sophomore Test Component
HIGH SCHOOL & BEYOND

Figure III-4

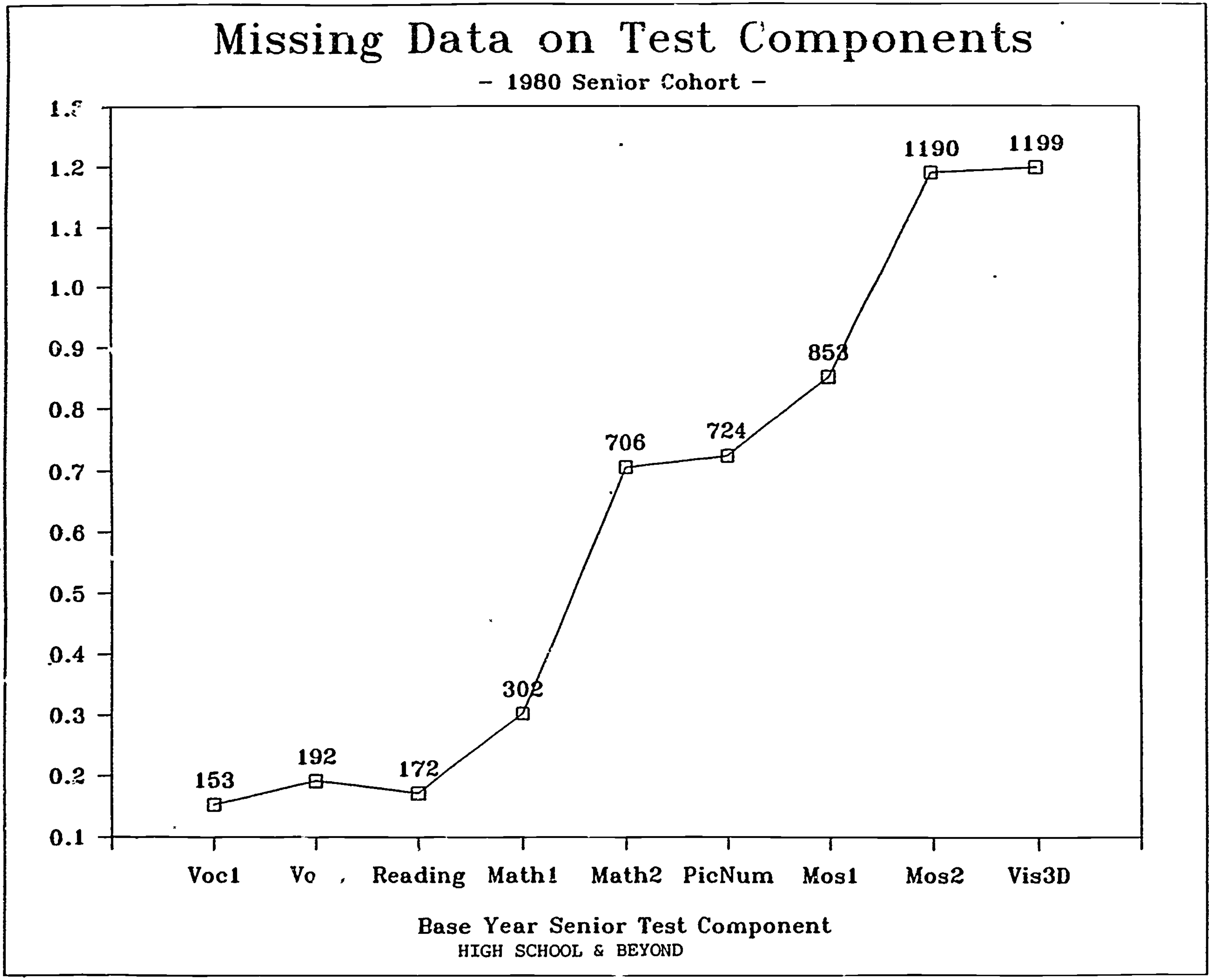
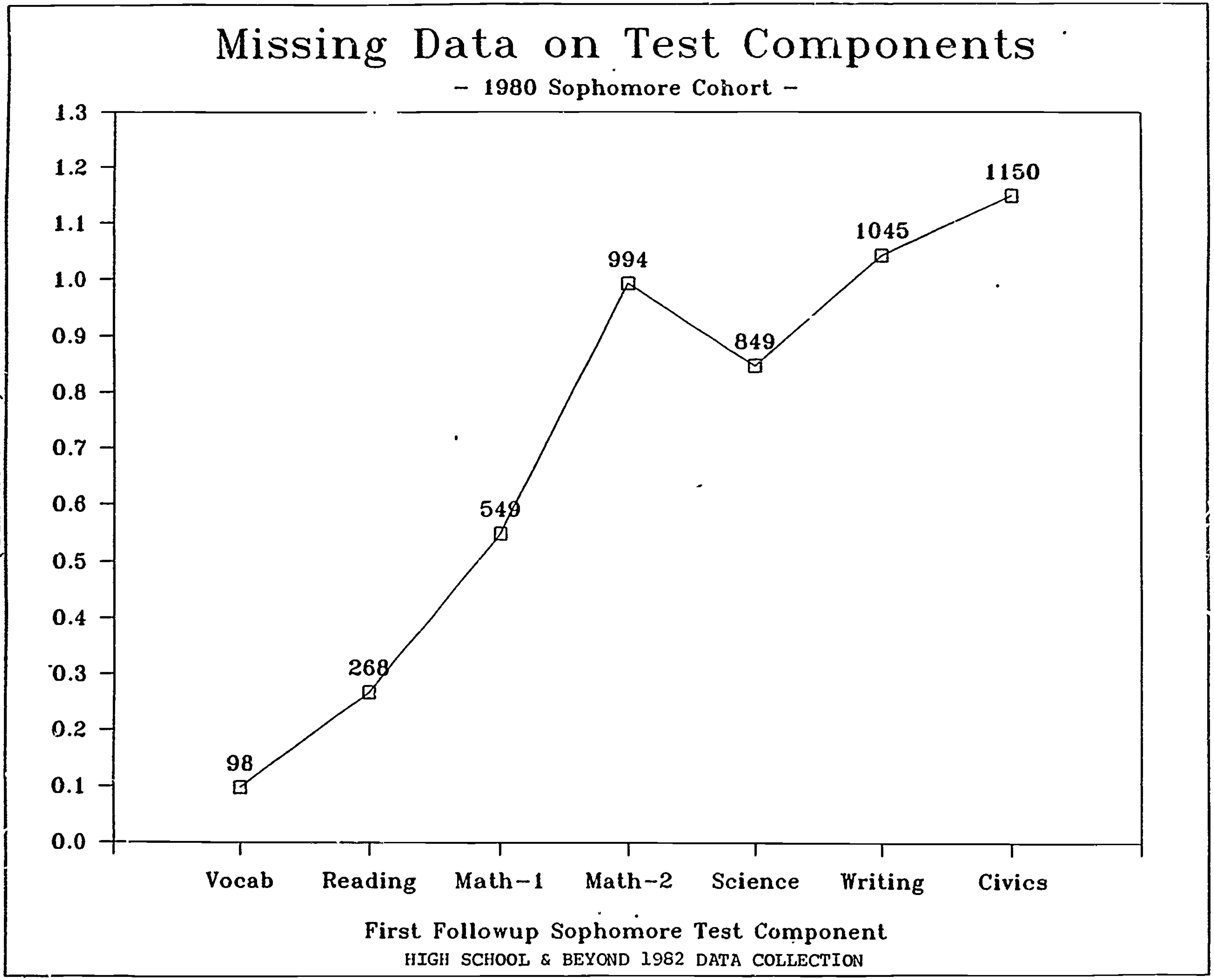

Missing Data on Test Components
- 1980 Senior Cohort -
Number Missing Component Score
(Thousands)
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
153
192
172
302
706
724
853
1190
1199
Voc1
Reading
Math1
Math2
PicNum
Mos1
Mos2
Vis3D
Base Year Senior Test Component
HIGH SCHOOL & BEYOND

Figure III-5



Missing Data on Test Components
- 1980 Sophomore Cohort -
Number Missing Component Score (Thousands)
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
98
268
549
994
849
1045
1150
Vocab
Reading
Math-1
Math-2
Science
Writing
Civics
First Followup Sophomore Test Component
HIGH SCHOOL & BEYOND 1982 DATA COLLECTION

REFERENCES.

Bradburn, N.M., L.J.Rips, and S.K.Shevell. 1987. Answering autobiographical questions: The impact of memory and inference on surveys. Science 236:157-161.

Frankel, M.R., L. Kohnke, D. Buonanno, and R. Tourangeau. 1981. High School and Beyond Base Year Sample Design Report. Chicago: National Opinion Research Center (NORC), University of Chicago.

Ingels, Steven J. et al. 1987. Field Test Report: National Education Longitudinal Study of 1988 (NELS:88). Report to NCES. Chicago: NORC, University of Chicago. ERIC ED 289 897

Ingels, Steven J. 1987. Evaluation of Item Nonresponse in the National Medical Care Utilization and Expenditure Survey. National Center for Health Statistics, NMCUES, Series A., Methodological Report No.3. DHHS Pub. No. 87-20003. Washington, D.C., U.S. Public Health Service: October 1987.

Ingels, Steven J., S.Y.Abraham, R.F.Carr, M.R. Frankel, K.A. Rasinski, and B.D. Spencer, NELS:88 Base Year Data File User's Manuals, forthcoming, Washington, DC: U.S. Department of Education, National Center for Education Statistics.

McCullagh, P. and J.A. Nelder. 1989. Generalized Linear Models. Routledge, Chapman and Hall.

National Center for Education Statistics. 1987 (March). Standards and Policies. Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement.

Platek, R. and G.B. Gray. 1983. Imputation Methodology in Incomplete Data in Sample Surveys. Vol II, Ed. by Madow, W.G., I. Olkin and D.B. Rubin. New York: Academic Press. pp. 239-333.

Rock, D.A, T.J. Hilton, J. Pollack, R.B. Ekstrom, M.E. Goertz. 1985. Psychometric Analysis of the NLS-72 and the High School and Beyond Test Batteries. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Rock, Donald and Judith Pollack. 1987. "The Cognitive Test Battery." In Ingels et al., 1987, pp. 168-183.

Rock, Donald and Judith Pollack. 1989. Psychometric Report for the NELS:88 Base Year Test Battery. Washington, D.C., U.S. Department of Education, National Center for Education Statistics

Schuman, H. and G. Kalton. 1986. Survey methods. In The Handbook of Social Psychology, 3rd ed. eds. G. Lindzey and E. Aronson, 635-697. Reading, MA: Addison-Wesley.

Spencer B.D., M.R. Frankel, S.J. Ingels, K.A. Rasinski, and R. Tourangeau. NELS:88 Base Year Sample Design Report. Forthcoming (1989): Chicago, NORC.

APPENDIX 1.

Exhibit A: NAEP exclusion criteria.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
CRITERIA FOR EXCLUDING STUDENT FROM THE ASSESSMENT

THE INTENT IS TO ASSESS ALL SELECTED STUDENTS. THEREFORE, ALL SAMPLED STUDENTS WHO ARE CAPABLE OF PARTICIPATING IN THE ASSESSMENT SHOULD BE ASSESSED

Some students may be incapable of participating meaningfully in the assessment because of limited English proficiency or a physical or mental handicap. The Local Administrator, with the advice of other staff members, may exclude from the assessment only those students who are incapable of taking the assessment because:

- The student is a native speaker of a language other than English and has been enrolled in an English-speaking public school (not including a bilingual education program) for less than two consecutive years;

OR

- The student is a special education student with a Individualized Education Plan (IEP) who is mainstreamed less than 50 percent of the time in academic subjects and the IEP team has determined that the student is unable to be assessed.

Students with limited English proficiency and students with IEP's should be assessed, if in the judgment of school staff, they are capable of taking the assessment. When there is doubt, include the student.

NOTE: It is expected that these more stringent criteria for exclusion will reduce the current excluded student rate on NAEP from around 5 percent to 4 - 4.5 percent (verbal communication, Jules Goodison, Educational Testing Service, to S. Ingels).

APPENDIX 1, continued.

Exhibit B: HS&B BASE YEAR INELIGIBILITY CRITERIA

HIGH SCHOOL & BEYOND INELIGIBILITY CRITERIA (1980 BASE YEAR)

A student was considered ineligible when that student:

- Transferred out of the selected school. (A transfer student was defined as a student who had left the school and whose records were requested for a new school).*
- Died.
- Would be unavailable until after August 31, 1980.
- Was listed on the roster in error.
- Had become a drop-out or lost student since he was selected. Such a student would have to have been out of school for 20 or more consecutive days and was not expected to return.**
- Was physically or mentally unable to participate in the survey. ***

*Transfers-out were not directly replaced. (However, as in NELS:88, all transfers-in were given a chance of selection into the sample). All other categories above in HS&B base year led to replacement by other students from the roster. (HS&B substituted students for the following cases: dropout, listed in error, language barrier, too ill [mentally, physically], in jail, unavailable entire field period, expelled, and deceased.) No substitution of students was done in the NELS:88 base year.

**Dropouts meet the 20 consecutive days criterion, are at least 16 years of age, and are not expected to return to school. Lost students are dropouts in all respects except that they are not 16 years of age.

***While this category was used to cover linguistic exclusion also, a Spanish language version of the questionnaire was provided so that students whose primary language competence was in Spanish would not be excluded. However, only 43 out of 30,030 1980 HS&B sophomores elected to use it.

APPENDIX 2. Sample freshening.

The NELS:88 Base Year is a representative sample of eighth grade students (and eighth grade schools) in the spring of 1988. As the 1988 eighth grade cohort advances through school and is followed up in 1990, most will be tenth graders though some will have skipped or been held back a grade and others will have dropped out, left the country, or died. The 1990 sample will be representative of tenth graders in the United States only if members of the spring 1990 tenth grade class who were not in eighth grade in the spring of 1988 also have some chance of selection into the study. By thus freshening the eighth grade cohort in 1990 with students who are recent immigrants, were held back or skipped grades so that they are not in the normal sequence, and others who were not eighth graders in spring of 1988, the NELS:88 tenth grade sample will be wholly comparable to the independently drawn tenth grade sample of High School and Beyond in 1980. In order for the NELS:88 1992 twelfth grade sample to be comparable to the nationally representative twelfth grade samples of NLS-72 in 1972, the High School and Beyond Sophomore Cohort in their senior year, 1982, and the HS&B Senior Cohort in 1980, freshening must be repeated for the NELS:88 Second Follow-Up. Note that this process will maintain the representativeness of the student sample. The school sample will not be nationally representative, nor will the NELS:88 tenth graders constitute a representative within-school sample of tenth graders (for example, at any given tenth grade school which contains a cluster of NELS:88 students, typically only one feeder school, the eighth grade school that was selected into NELS:88, will be represented). In order to increase the generalizability of school-level NELS:88 data in 1990 and 1992, NORC and NCES are exploring statistical procedures that might permit the simulation and effective approximation of selection probabilities for the NELS:88 high schools, thus permitting a school weight to be attached to each school in the 1990 and 1992 samples. In addition, in order to study school effects, we are exploring the possibility of increasing the cluster size and representativeness of the within-school sample within a subsample of schools. If school selection probabilities can be successfully approximated, then this subsample of schools would constitute a true sample from which school effects conclusions could be generalized. It is too early, however, to say whether either the attempt to estimate school selection probabilities, or the attempt to implement a school effects supplement, will prove possible in coming months. In any case, however, the representativeness of the student sample at the national level will be guaranteed by sample freshening.

A technique for sample freshening in NELS:88 was developed by NORC Senior Statistical Scientist Martin R. Frankel. This linking technique for obtaining a representative sample of tenth grade students from a 1988 sample of eighth graders draws on L. Kish's half-open interval procedure. Frankel's original suggestion is quoted in Spencer* (1987, p.124):

> The eighth grade student population will mostly, but not entirely, be enrolled in tenth grade in 1990. To keep the description simple we will assume that all schools enrolling tenth grade students in 1990 enroll some tenth graders who were eighth graders in 1988. First, pick a subsample (or all) of the Base Year sample who are enrolled in tenth

grade in 1990. Consider any single school enrolling at least one of those students and list all its tenth grade students in a circular list (for example, alphabetically from A to Z followed by A again). Identify each student in the list who was in the Base Year sample and go down the list to the next student. If that student was an eighth grader in 1988, stop; otherwise include that student in the tenth grade sample and go down the list to the next student. If that student was an eighth grader in 1988, stop, and otherwise include that student in the tenth grade sample and go down the list to the next student, and so on. Repeat this procedure for all sampled 1988 eight grade students who were in tenth grade in 1990 and for all schools. To determine sample weights, note that the selection probability for a newly selected tenth grader is equal to that for the sampled eighth grader closest up on the list.

*Bruce D. Spencer, Sampling Problems in Merging a Cross-Sectional and a Longitudinal Program." (1987). In The National Assessment of Educational Progress and the Longitudinal Studies Program: Together or Apart?--Report of a Planning Conference, December 11, 1986. Prepared by George H. Brown and E.M. Faupel. Washington, D.C.: National Center for Education Statistics, CS 87-446.