

DOCUMENT RESUME

ED 234 640

FL 013 962

AUTHOR So, Alvin Y.  
 TITLE The High School and Beyond Data Set: Its Relevance for Bilingual Education Research.  
 INSTITUTION National Center for Bilingual Research, Los Alamitos, Calif.  
 SPONS AGENCY National Inst. of Education (ED), Washington, DC.  
 REPORT NO NCBR-TN-5  
 PUB DATE 22 Oct 82  
 NOTE 13p.; For related document, see FL 013 966.  
 AVAILABLE FROM National Center for Bilingual Research, 4665 Lampson Ave., Los Alamitos, CA 90720  
 PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.  
 DESCRIPTORS Academic Aspiration; \*Bilingual Education; Data; \*High School Students; Language Usage; \*National Surveys; Secondary Education; Student Characteristics  
 IDENTIFIERS \*High School and Beyond (NCES)

ABSTRACT

Preliminary data from the High School and Beyond (HSB) research study are described in order to assist bilingual education researchers in understanding what information is available. The HSB project design included a highly stratified national probability sample encompassing 30,000 sophomores and 28,000 seniors enrolled in 1,015 public and private high schools. The study seeks to observe the educational and occupational plans and activities of high school students as they pass through the American educational system. The nature of the various data files is described including files on students, languages, schools, teachers' comments, parents, tests, twins, and friends. For example, the most important file, the student file, contains responses from each student to extensive questionnaires and various cognitive tests. The language file contains information distinguishing childhood language status from present language status, language usage at home versus language usage outside of the home, and information describing experience with bilingual education. The constraints of the sample that limit its generalizability are discussed. It is concluded that, keeping sample constraints in mind, the HSB data provide an extremely valuable resource for bilingual education researchers. (RW)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

ED234640

THE HIGH SCHOOL AND BEYOND DATA SET:  
ITS RELEVANCE FOR BILINGUAL EDUCATION RESEARCH

Alvin Y. So

October 22, 1982

TN-5

U.S. DEPARTMENT OF EDUCATION  
NATIONAL INSTITUTE OF EDUCATION  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it. Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

**NCBR**

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

FL013962

National Center for Bilingual Research, 4665 Lampson Avenue, Los Alamitos, CA 90720, telephone (213) 598-0481



The material in this publication was prepared under Cooperative Agreement 00-CA-80-0001 from the National Institute of Education, Department of Education. Its contents do not necessarily reflect the views of the National Institute of Education or of any other agency of the United States government, and no official endorsement should be inferred.

## ABSTRACT

Preliminary analysis of the High School and Beyond (HS&B) data set reveals that, despite its sample constraints, it will be an invaluable set of resources for researchers of bilingualism and bilingual education. The special inclusion of the Hispanic population, the largest language minority in the U.S., will enable researchers to carry out detailed analyses on that population. In addition, the various files of the HS&B data set includes needed variables to test the validity of many of the heated arguments surrounding bilingual education.

## THE HIGH SCHOOL AND BEYOND DATA SET: ITS RELEVANCE FOR BILINGUAL EDUCATION RESEARCH

Alvin Y. So

### Introduction

Funded by the National Center for Education Statistics (NCES) and conducted by the National Opinion Research Center (NORC), the High School and Beyond (HS&B) data set was the first wave of a national longitudinal study of the cohorts of high school students in the United States in 1980. The HS&B project design included a highly stratified national probability sample of over 11,000 high schools with 36 seniors and 36 sophomores per school. In those schools with fewer than 36 seniors or sophomores, all eligible students were included in the sample. Cooperation from both schools and students was excellent. The overall response rate for schools was 91% and for students, 84%. Over 30,000 sophomores and 28,000 seniors enrolled in 1,015 public and private high schools across the nation participated in this study. The HS&B sample represents the nation's 10th and 12 grade populations, totaling about 3.8 million sophomores and 3 million seniors in more than 21,000 schools in spring, 1980 (Peng et al., 1981, p. ix; NORC, 1980a).

As a large-scale, longitudinal survey, the primary purpose of the HS&B project is to observe the educational and occupational plans and activities of young people as they pass through the American educational system and assume their adult roles (Peng et al., 1981, p. ix). Because of its excellent sample and questionnaire design, however, the HS&B project actually has collected much more data than required for its original purpose. Page (1981, pp. 22-23) describes it as a "priceless national resource. . . . It is an extraordinarily far-sighted project, the richest resource for research and policy analysis we have had." Subsequently, many well-known researchers such as James Coleman (1981) have utilized the HS&B data set to generate publications that have important policy implications. Recently, many educational journals have devoted

their entire issue to a policy report based on this data set (see, for example, Harvard Educational Review, 1981; Sociology of Education, 1982).

Despite its rich data, however, the HS&B data set has still not caught the attention of researchers of bilingual education. Except for the pioneer study by Nielsen and Fernandez (1981) on the achievements of Hispanic students, no work has been done from the perspective of bilingualism or bilingual education utilizing this rich resource.

Because of the lack of familiarity of bilingual education researchers with the HS&B data set, we have conducted preliminary analyses of the data and are documenting the data set in this technical note to introduce the data set to our fellow research workers in bilingual education. In what follows we shall first describe the nature of various data files contained in the HS&B data set. Then we shall point out the reason why this data set is particularly useful to bilingual education researchers, noting for the reader the constraint imposed by this data set in carrying out bilingual education research.

### The Data Files and the Variables

In order to collect data from as many different resources as possible, the HS&B project distributed several sets of questionnaires to various individuals. The data collected were then stored in different computer files, as presented in Table 1. We shall briefly describe each of these files in the following sections.

The student file. The most important file in the HS&B data set, the student file contains responses from each student in the sample to a fairly extensive questionnaire and to various cognitive tests. Consequently, this file contains responses from all the 58,000 students in the HS&B sample and includes as many as 638 variables. A summary listing of the variables in this file is as follows:

- High School Experience Variables (curriculum placement, course taken, grades and homework, vocational training, students' opinion of the school)

- Activities Outside of School Variables (working for pay, organized group activities, other leisure activities)
- Values and Attitudes Variables (life goals, factors in educational and occupational choices, national services)
- Plans of High School Seniors Variables (short-range plans, long-range plans)
- College Plans Variables (criteria for choosing a college, financial aid, expected field of study)
- Achievement Tests Variables (vocabulary, reading, mathematics, picture-number, mosaic comparison, visualization in three dimensions)

Table 1: A List of the Data Files in the HS & B Data Set

Name of the File	Number of Cases in the File	Number of Variables in the File
The Student File	58,000 students	638
The Language File	11,000 students with non-English language experience	42
The School File	988 schools	237
The Teachers' Comment File	143,000 teacher observations	30
The Parent File	7,000 parents	307
The Test File	53,000 students	248
The Twin File	500 twins	640
The Friend File	36,000 one-way friendship linkages	not specified

The language file: If a student reported some non-English language experience either during childhood or at the time of the survey, the student was requested to complete an additional set of questionnaires on language experience. About 11,000 out of a total of 58,000 students answered the language questionnaire; their responses were included in the

language file. A summary listing of the 42 variables in this file is as follows:

- Language Status as a Child
- Present Home Language Variables
- Self-Assessed English and Other Language Proficiency (understanding, spoken, reading, writing)
- Present Language Usage (at home, at school, at work, at store)
- Experience with Bilingual Medium of Instruction in Grades 1-6, 7-9, 10-12
- Courses Taken (in English as a Second Language, reading/writing, math/science courses taught in other language, ancestry history)

Since this file will be of most interest to bilingual education researchers, a brief description of the sample characteristics of this file is presented in Table 2. This language file contains responses from 5,120 Hispanics, 3,763 Whites, 663 Asians, 203 American Indians, and 162 Blacks; about 1,100 students in this file did not answer the question on either descent or mother tongue. Of all the ethnic groups in this language file, the highest percentage (67%) of non-English mother tongue students were Asians, followed closely by Hispanics (62%); by American Indians (39%), and by Whites (20%); Blacks turned out to be lowest percentage (only 8%) of non-English mother tongue students.

Table 2: Sample Characteristics of the Language File

Mother Tongue	Ethnic Groups					Total
	Hispanics	Whites	Asians	Am. Indians	Blacks	
English	38%	80%	34%	62%	92%	55%
Spanish	61%	2%	2%	3%	4%	32%
Other Language	1%	18%	65%	36%	4%	13%
Total %	100%	100%	101%	101%	100%	100%
(N)	(5,120)	(3,763)	(663)	(203)	(162)	(9,911)

The school file. The administrator in each selected school in the HS&B sample was requested to complete a questionnaire about the school; their responses are included in this school file. This file provides information about the social context in which the students receive their high school education. All in all, 988 school administrators responded to questions containing some 237 variables. A summary list of the variables are:

- School Facilities Variables (library volumes, indoor lounge, departmental office, student cafeteria)
- School Educational Characteristics (highest/lowest grade offered, total membership, length of school year, average daily attendance, number of graduates)
- School Ethnic Composition Variables (percentage of American Indian, Asian, Hispanic, Black, White students and faculty)
- School Social Environment Variables (student absenteeism, cutting classes, parents' lack of interest, teacher absenteeism, robbery, drugs, rape, vandalism)
- School Financial Situation Variables (per-student expenditure, percentage of funds from tuition, from fund-raising, from religious subsidy, annual tuition, legal ownership)
- Teacher Characteristics (percentage female, percentage MA, average pay, salary steps, teaching experience)
- Language Courses taught (Spanish, German, French, Black Studies cultural courses, bilingual program, ESL courses, courses taught in mother tongue)

The teachers' comments file. Teachers in each selected school in the HS&B sample were asked to make comments on students identified in the sample. About 14,000 teachers from 611 schools responded on about 17,000 students. Since a teacher could make comments on one or more students, there were a total of about 143,000 teacher observations in this file. A partial list of the 30 variables in this file is as follows:

- Classes Taught by Teacher (English, art, history, etc.)
- Social Background of the Teacher (sex, ethnicity)
- Teacher's Knowledge of the Student (had student in class, know student, know parent)



- Evaluation of Student's Performance (student working up to potential, will probably go to college, seems to dislike school)
- Comments on Student's Social Traits (seems popular with others, emotional handicaps, self-discipline to hold a job)

The parent file. About 7,000 parents of the students in the HS&B sample were selected to complete another set of questionnaires containing their views on high school education. A list of the 307 variables in this file is as follows:

- Parent's Social Background Variables (sex, ethnicity, education, occupation, industry, language status, social mobility)
- Parent's Communication with Students (talk to students in grades 6-7, 8-9, 10-11, 12)
- Parent's Expectation of Students' Educational and Occupational Achievements
- Parent's Ability to Finance College Education
- Parent's Actual Involvement in College Planning (talking to counselors, reading pamphlets, talking to other parents)

Finally, there are the test file, the friend file, and the twin file which include a battery of cognitive tests, friendship linkages and information on twins, respectively. Since these three files may be of less interest to bilingual education researchers, we shall not review them here. Interested readers can consult the codebooks or news releases for further details (NCES, 1982a, 1982c; NORC, 1980a, 1980b, 1980c).

#### The Relevance of the HS&B Data Set to Bilingual Education Research

The HS&B data set is particularly useful to researchers in bilingualism and bilingual education because of its excellent language file. According to Nielsen and Fernandez (1981, p. 3), the language file contains a language questionnaire that is even superior in quality to that in the 1976 Survey of Income and Education national data set.

First, the language file distinguishes childhood language status from the present language status, thus permitting researchers to study the rate of language shift in the present generation of high school students. Secondly, the language file also distinguishes language usage at home from usage outside the home, and distinguishes oral proficiency (speaking, listening) from literacy (reading, writing). These finer distinctions enable researchers to study in more detail the actual patterns of language shift in these four important language domains. Third, the language file includes information on experience with a bilingual medium of education and on types of language courses taken in schools. This kind of language information allows researchers to classify types of bilingual education programs and to investigate their differential impacts on language shifts.

In addition, when the language file is merged with other files in the HS&B data set, the newly merged file provides important data that can open up new frontiers in bilingual education research. For instance, the merged language-student file will allow researchers to study the social background of language minority students, their experience in the U.S. high schools, and their educational achievement in comparison with non-language-minority youths.

Another example is the merged language-school file, which will enable researchers to study language minority youths from a holistic perspective. The new language-school file will tell us, for example, which type of schools do most language minority students attend, what is the ethnic composition of the students and the social environment in those schools, and what kinds of language courses are offered.

In addition to the rich number of variables it contains, the HS&B data set is also valuable to bilingual education in that it includes information on the largest language minority in the U.S., i.e., Hispanics. Rarely has a national survey on high school education paid sufficient attention to the issues facing the Hispanic language minority. Thus, the HS&B data set may be the first national project that aims to include adequate Hispanic respondents in its sample.

To achieve this aim, the HS&B project designed a special strata of 136 Hispanic schools. Further, out of its 58,000 student sample, the HS&B data set included about 6,700 Hispanic students. Since over-sampling would affect the representativeness of the HS&B sample, weights were assigned to each student in the sample in order to balance out the over-sampling effect. Consequently, weights were calculated to reflect differential probabilities of sample selection and to adjust for nonresponse. In this respect, the HS&B data set remains a nationally representative study in which its sample characteristics can be used to deduce the U.S. student population.

#### Constraints on the Sample which Limit Generalizability

This section of the technical note points to the sample constraints of the HS&B data set for conducting bilingual education research. First of all, 8,278 students, or about 12% of the originally targeted 69,662 student sample, were absent on the day the HS&B survey was conducted (NORC, 1980a, Table 1). Since this represents quite a large number of students, it cannot be assumed that all the absentees were sick or were absent for family reasons. It is highly conceivable that many of these absentee students were from language minority backgrounds, but there is no way to really estimate the number of absentee language minority students. If this assumption is correct, then the HS&B data set has already discarded quite a large number of language minority background students from the sample.

The second sample constraint, which follows the logic of the above argument but has more serious consequences, is the high drop out rate for language minority students. Chan (in press) points out that the drop out rate for limited-English or non-English speaking children is about three to four times the rate for English-speaking students. Similarly, Waggoner (1981, p. 41) reveals that language minority students are less than half as likely as people with English language backgrounds to have completed high school or to have attended college. Similarly, Nielsen and Fernandez (1981, p. 14) also suggest that among

Hispanic dropouts, 60% had left school before grade 10. These studies point to the fact that many language minority students with low reading achievement have dropped out of school before grade 10. Consequently, the HS&B data set at best includes only those students who are talented or determined enough to survive through high school beyond the 10th grade.

Due to the above filtering processes of absenteeism and dropping out, the third sample constraint which necessarily follows is that there is the conspicuous absence of non-English speaking language minority students in the HS&B data set. A simple fact is that if a student really is non-English speaking, that student could not make it to grade 10 and show up on the HS&B survey day. Consequently, when a student is asked for his/her self-assessed English ability on the HS&B questionnaire, almost no one in the sample replied that he/she did not understand English. Indeed, one has to understand what is written on the English HS&B questionnaire at least well enough in order to circle the right answer "no English ability at all." Consequently, only 56 out of 58,000 students answered the questionnaire in Spanish. And of these 56 students, only 11 showed up in the language file, a fact that continues to puzzle us.

It is hard to assess what impacts these sample constraints might have on bilingual education research. We can speculate that results from analyses of the HS&B data set might tend to overestimate language shift towards English monolingualism, and to underestimate the educational disadvantages facing language minority students because of a large number of students who were either absent from school on the day of the survey or who were drop outs. Keeping these sample constraints in mind, however, the HS&B data set will prove to be an extremely valuable resource for researchers in bilingual education.

References

- Chan, K. S. Limited English speaking, handicapped, and poor: Triple threat in childhood. In M. Chu-Chang (Ed.), Asian- and Pacific-American Perspectives in Bilingual Education: Comparative Research. New York: Teachers College Press, in press.
- Coleman, J. S.; Hoffer, T.; & Kilgore, S. Public and Private Schools: Report to the National Center for Education Statistics. Chicago: National Opinion Research Center, 1981.
- National Center for Education Statistics. High School and Beyond, Users Manual for Teacher Comment File. Washington, D.C.: National Center for Education Statistics, Longitudinal Studies Branch, 1982a.
- National Center for Education Statistics. Hispanic students in American high schools. NCES Bulletin, July, 1982b.
- National Center for Education Statistics. Field file for High School and Beyond Data now available. NCES Announcement, September, 1982c.
- National Opinion Research Center. High School and Beyond: Information for Users. Base Year 1980. Report to the National Center for Educational Statistics under contract #300-78-0208. Chicago: National Opinion Research Center, 1980a.
- National Opinion Research Center. High School and Beyond: Language File Codebook. Chicago: National Opinion Research Center, 1980b.
- National Opinion Research Center. High School and Beyond: School Questionnaire Codebook, Version I. Chicago: National Opinion Research Center, 1981.
- Nielsen, F., & Fernandez, R. M. Achievement and Hispanic students in American high schools. Report to the National Center for Education Statistics under contract #300-78-0208. Chicago: National Opinion Research Center, 1981.
- Page, E. B. The media, technical analysis, and the data feast. Educational Research, 1981, 10, 21-23.
- Peng, S. S.; Feters, W. B.; & Kilstad, A. J. High School and Beyond: A national longitudinal study for the 1980s. A capsule description of high school students. Washington, D.C.: National Center for Education Statistics, 1981.
- Report analysis: Public and private schools. Harvard Educational Review, 1981, 51, 481-545.
- Special issue on public and private schools. Sociology of Education, 1982, 55, 63-182.
- Waggoner, D. Educational attainment of language minorities in the United States. NABE Journal, 1981, 6, 41-53.