

High School and Beyond Data Set

The high school and Beyond data set (hsb) is from a large-scale longitudinal study conducted by the National Opinion Research Center (1980) under contract with the National Center for Education Statistics. The students in the study are a nationally representative sample of $n = 600$ high school seniors with observations on 15 variables.

The data are in a file named “hsb.dat”, which will be made available to you. In this file, there are 600 lines (one for each subject) and 15 columns (one for each variables). Brief descriptions of the variables in the order in which they are in the file are given below.

Variable Number	Variable Name	Variable Label	Coded Responses
1	ID	Student identification number	(3-digit consecutive numbers)
2	SEX	Sex	1 = Male 2 = Female
3	RACE	Race or ethnicity	1 = Hispanic 2 = Asian 3 = Black 4 = White
4	SES	Socioeconomic status	1 = Low 2 = Medium 3 = High
5	SCTYP	School type	1 = Public 2 = Private
6	HSP	High school program	1 = General 2 = Academic preparatory 3 = Vocational/technical

Variable Number	Variable Name	Variable Label	Coded Responses
7	LOCUS	Locus of control	Standardized to mean 0 & SD of 1
8	CONCPT	Self-concept	Standardized to mean 0 & SD of 1
9	MOT	Motivation	Average of 3 motivation items
10	CAR	Career choice	1 = Clerical 2 = Craftsman 3 = Farmer 4 = Homemaker 5 = Laborer 6 = Manager 7 = Military 8 = Operative 9 = Professional 1 ¹ 10 = Professional 2 ² 11 = Proprietor 12 = Protective 13 = Sales 14 = School 15 = Service 16 = Technical 17 = Not working
11	RDG	Reading T-score	standardized to mean 50 and SD 10
12	WRTG	Writing T-score	standardized to mean 50 and SD 10
13	MATH	Math T-score	standardized to mean 50 and SD 10
14	SCI	Science T-score	standardized to mean 50 and SD 10
15	CIV	Civics T-score	standardized to mean 50 and SD 10

1 accountant, nurse, engineer, librarian, writer, social worker, athlete, politician, etc.

2 clergyman, dentist, physician, lawyer, scientist, college teacher, etc.