

ICPSR 34437

**Gates Millennium Scholars  
Program (GMS) Cohort 2,  
2001-2006**

*Bill and Melinda Gates Foundation*

Codebook for Freshmen Cohort 2 (2002)  
College Board Data

Inter-university Consortium for  
Political and Social Research  
P.O. Box 1248  
Ann Arbor, Michigan 48106  
[www.icpsr.umich.edu](http://www.icpsr.umich.edu)

# Terms of Use

The terms of use for this study can be found at:  
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/34437/terms>

## Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

### NOTICE

#### WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

**ICPSR PROCESSING NOTE FOR # 34437**  
*Gates Millennium Scholars Program (GMS) Cohort 2, 2001-2006*  
**Public use data**

**1) Confidentiality:**

- In the public use data, ICPSR has blanked the all variables with open ended responses to preserve respondent anonymity.
- ICPSR has recoded numeric variables with sensitive information using top and bottom coding methods to protect respondent confidentiality.
- All IPEDS codes were blanked in the public version of the data.
- For access to the restricted use data, the user must sign the ICPSR Restricted Data Use Agreement.
- Values that were deemed as indirect identifiers (such as age, income, GPA) ICPSR used a series of data coarsening techniques to protect for confidentiality. More specifically, GPA variables were grouped by quartiles, while age and income were top-coded.

**2) Question Text:**

- Question text was added to the data based on the original data collection instrument deposited by the Principal Investigator.

**3) Value labels:**

- Value labels were standardized based on the original data collection instrument deposited by the Principal Investigator.
- Any spelling errors found in the value labels were corrected during ICPSR processing.

**4) Missing Values:**

- Missing values were standardized as per the study's convention. Values "-4" and "-5" were recoded or labeled as "Logical Skip" and "Missing Data" respectively.

**5) Unknown code:**

- Any unknown codes were given the value label "Unknown Code – ICPSR".

**6) Special Considerations**

- Credits, grades and GPA were reported in a wide variety of ways in the data; no effort has been made to regularize or convert grades, GPA or credits. Users of this study should use caution when utilizing these variables in their research.

## Data Completeness Report

Notes: (1) Variables are individually listed only if they have greater than 5% missing data. These variables are listed under the appropriate percentage category in the order in which they appear in the data file. (2) The Data Completeness Report only captures information about system missing or other values that are declared missing. Codes that have a label implying that they are missing but that are not declared missing values are not reflected in this report. Data users should consult the codebook for more specific information about missing values. (3) Some variables that have 100% missing data may have been blanked by ICPSR to protect respondent confidentiality. Data users should consult the codebook for more specific information about blanked variables. (4) Data do not contain skip patterns or skip patterns are not reflected in the data as coded.

**Table 1: Distribution of Variables by Percentage of Missing Values--Freshmen Cohort 2 (2002)**

### College Board Data

Variable Name and Label (Total Cases = 2067 )	Percent of Cases with Missing Values
84.6% ( 22 of 26 variables)	have 0% Missing Values
0.0% ( 0 of 26 variables)	have 0% - 1% Missing Values
0.0% ( 0 of 26 variables)	have 1% - 3% Missing Values
0.0% ( 0 of 26 variables)	have 3% - 5% Missing Values
0.0% ( 0 of 26 variables)	have 5% - 10% Missing Values
7.7% ( 2 of 26 variables)	have 10% - 20% Missing Values
PUBLICHS	Public High School 12.0%
GPA	High School GPA 14.1%
3.8% ( 1 of 26 variables)	have 20% - 40% Missing Values
CLASRANK	High School Class Rank 26.8%
3.8% ( 1 of 26 variables)	have 40% - 99% Missing Values
GRADTYPE	Graduation Type 85.5%
0.0% ( 0 of 26 variables)	have 100% missing values

# **ICPSR 34437**

## **Gates Millennium Scholars Program (GMS) Cohort 2, 2001-2006**

### **Variable Description and Frequencies**

**Note: Frequencies displayed for the variables are not weighted. They are purely descriptive and may not be representative of the study population. Please review any sampling or weighting information available with the study.**

**Summary statistics (minimum, maximum, mean, median, and standard deviation) may not be available for every variable in the codebook. Conversely, a listing of frequencies in table format may not be present for every variable in the codebook either. However, all variables in the dataset are present and display sufficient information about each variable. These decisions are made intentionally and are at the discretion of the archive producing this codebook.**

## Gates Millennium Scholars Program (GMS) Cohort 2, 2001-2006

### Freshmen Cohort 2 (2002) College Board Data

SU_ID	Respondent ID
-------	---------------

Location: 1-8 (width: 8; decimal: 0)  
 Variable Type: character  
 Question: Respondent ID  
 Based upon 2067 valid cases out of 2067 total cases.

PL_2	Preload Population Type
------	-------------------------

Location: 9-10 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Question: Preload Population Type

Value	Label	Unweighted Frequency	%
1	2000 Freshman recipients	0	0.0 %
2	2000 Continuing undergraduate recipients	0	0.0 %
3	2000 Graduate recipients	0	0.0 %
4	2001 Freshman recipients	1070	51.8 %
5	2000 Freshman non-recipients	0	0.0 %
6	2000 Continuing undergraduate non-recipients	0	0.0 %
7	2000 Graduate non-recipients	0	0.0 %
8	2001 Freshman non-recipients	997	48.2 %
9	2002 Freshman recipients	0	0.0 %
10	2002 Freshman non-recipients	0	0.0 %
11	2004 Freshman recipients	0	0.0 %
12	2004 Freshman non-recipients	0	0.0 %

Based upon 2067 valid cases out of 2067 total cases.

- Mean: 5.93
- Median: 4.00
- Mode: 4.00
- Minimum: 4.00
- Maximum: 8.00
- Standard Deviation: 2.00

RACE_DB	Preload Race
---------	--------------

Location: 11-12 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Question: Preload Race

- Study 34437 -

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
1	African Americans	747	36.1 %
2	American Indians	264	12.8 %
3	Asian/Pacific Islanders	408	19.7 %
4	Hispanic Americans	648	31.3 %

Based upon 2067 valid cases out of 2067 total cases.

- Mean: 2.46
- Median: 3.00
- Mode: 1.00
- Minimum: 1.00
- Maximum: 4.00
- Standard Deviation: 1.26

**SEX\_DB Preload Sex/Gender**

Location: 13-14 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Question: Preload Sex/Gender

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
1	Male	643	31.1 %
2	Female	1424	68.9 %
3	Unknown	0	0.0 %

Based upon 2067 valid cases out of 2067 total cases.

- Mean: 1.69
- Median: 2.00
- Mode: 2.00
- Minimum: 1.00
- Maximum: 2.00
- Standard Deviation: 0.46

**PELL\_DB Preload Pell Grant Eligibility**

Location: 15-16 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Question: Preload Pell Grant Eligibility

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
1	Pell-Eligible	1072	51.9 %
2	Pell-Ineligible	345	16.7 %
3	Pell Eligible Unknown	650	31.4 %

- Study 34437 -

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
4	Not Applicable (i.e Graduate student)	0	0.0 %

Based upon 2067 valid cases out of 2067 total cases.

- Mean: 1.80
- Median: 1.00
- Mode: 1.00
- Minimum: 1.00
- Maximum: 3.00
- Standard Deviation: 0.89

**SEL\_NSEL Preload Select or No-select**

Location: 17-18 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Question: Preload Select or No-select

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
1	Select	1813	87.7 %
2	Non-Select	254	12.3 %

Based upon 2067 valid cases out of 2067 total cases.

- Mean: 1.12
- Median: 1.00
- Mode: 1.00
- Minimum: 1.00
- Maximum: 2.00
- Standard Deviation: 0.33

**REC\_NREC Preload Scholar or Non-Recipient**

Location: 19-20 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Question: Preload Scholar or Non-Recipient

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
1	Scholar	1070	51.8 %
2	Non-Recipient	997	48.2 %

Based upon 2067 valid cases out of 2067 total cases.

- Mean: 1.48
- Median: 1.00
- Mode: 1.00



- Study 34437 -

- Minimum: 1.00
- Maximum: 2.00
- Standard Deviation: 0.50

**SCH\_YR School Year**

Location: 21-22 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Question: School Year

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
1	2000-2001	0	0.0 %
2	2001-2002	2067	100.0 %
3	2002-2003	0	0.0 %
4	2003-2004	0	0.0 %
5	2004-2005	0	0.0 %

Based upon 2067 valid cases out of 2067 total cases.

- Mean: 2.00
- Median: 2.00
- Mode: 2.00
- Minimum: 2.00
- Maximum: 2.00
- Standard Deviation: 0.00

**ORIG\_COLL Original Uncorrected College Board ID**

Location: 23-62 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: Original Uncorrected College Board ID

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**CBID Corrected College Board ID**

Location: 63-102 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: Corrected College Board ID

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

- Study 34437 -

**NCESID** **NCES School ID**

Location: 103-142 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: NCES School ID

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**PUBLICHS** **Public High School**

Location: 143-144 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Range of Missing Values (M): -5  
 Question: Public High School

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
0	Private, Foreign, Competitive Entry, Home School, or Foreign	203	9.8 %
1	Regular Public or DOD School	1615	78.1 %
-5 (M)	Missing Data	249	12.0 %

Based upon 1818 valid cases out of 2067 total cases.

- Mean: 0.89
- Median: 1.00
- Mode: 1.00
- Minimum: 0.00
- Maximum: 1.00
- Standard Deviation: 0.32

**CBNAME** **College Board High School Name**

Location: 145-184 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: College Board High School Name

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**CBCITY** **College Board High School City**

Location: 185-224 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: College Board High School City

- Study 34437 -

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**CBSTATE**                      **College Board High School State**

Location:                      225-264 (width: 40; decimal: 0)  
 Variable Type:              character  
 Question:                      College Board High School State

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**CBZIP**                              **College Board High School Zipcode**

Location:                      265-304 (width: 40; decimal: 0)  
 Variable Type:              character  
 Question:                      College Board High School Zipcode

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**CBFORADD**                      **College Board Foreign County High School Address**

Location:                      305-344 (width: 40; decimal: 0)  
 Variable Type:              character  
 Question:                      College Board Foreign County High School Address

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**CLASRANK**                      **High School Class Rank**

Location:                      345-346 (width: 2; decimal: 0)  
 Variable Type:              numeric  
 Range of Missing Values (M): -5  
 Question:                      High School Class Rank

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
-4	Logical Skip	0	0.0 %

- Study 34437 -

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
1	1st Quartile Class Rank - ICPSR ADDED	394	19.1 %
2	2nd Quartile Class Rank - ICPSR ADDED	388	18.8 %
3	3rd Quartile Class Rank - ICPSR ADDED	353	17.1 %
4	4th Quartile Class Rank - ICPSR ADDED	378	18.3 %
-5 (M)	Missing Data	554	26.8 %

Based upon 1513 valid cases out of 2067 total cases.

- Mean: 2.47
- Median: 2.00
- Mode: 1.00
- Minimum: 1.00
- Maximum: 4.00
- Standard Deviation: 1.13

**CLASSIZE High School Class Size**

Location: 347-386 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: High School Class Size

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**GPA High School GPA**

Location: 387-388 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Range of Missing Values (M): -5  
 Question: High School GPA

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
-4	Logical Skip	0	0.0 %
1	1st Quartile GPA - ICPSR ADDED	446	21.6 %
2	2nd Quartile GPA - ICPSR ADDED	438	21.2 %
3	3rd Quartile GPA - ICPSR ADDED	393	19.0 %
4	4th Quartile GPA - ICPSR ADDED	498	24.1 %
-5 (M)	Missing Data	292	14.1 %

Based upon 1775 valid cases out of 2067 total cases.

- Mean: 2.53

- Study 34437 -

- Median: 3.00
- Mode: 4.00
- Minimum: 1.00
- Maximum: 4.00
- Standard Deviation: 1.15

**GRADTYPE                      Graduation Type**

Location: 389-390 (width: 2; decimal: 0)  
 Variable Type: numeric  
 Range of Missing Values (M): -5  
 Question: Graduation Type

<i>Value</i>	<i>Label</i>	<i>Unweighted Frequency</i>	<i>%</i>
3	-	276	13.4 %
7	-	23	1.1 %
-5 (M)	Missing Data	1768	85.5 %

Based upon 299 valid cases out of 2067 total cases.

- Mean: 3.31
- Median: 3.00
- Mode: 3.00
- Minimum: 3.00
- Maximum: 7.00
- Standard Deviation: 1.07

**GEDSCORE                      GED Score**

Location: 391-430 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: GED Score

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**GRADREC                      Best guess at which HS record is the last HS attended (one per student)**

Location: 431-470 (width: 40; decimal: 0)  
 Variable Type: character  
 Question: Best guess at which HS record is the last HS attended (one per student)

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

- Study 34437 -

**BEGINDATE                      Begin Date**

Location:                      471-510 (width: 40; decimal: 0)  
 Variable Type:              character  
 Question:                      Begin Date

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**ENDDATE                         End Date**

Location:                      511-550 (width: 40; decimal: 0)  
 Variable Type:              character  
 Question:                      End Date

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.

**GRADDATE                       High School Graduation Date**

Location:                      551-590 (width: 40; decimal: 0)  
 Variable Type:              character  
 Question:                      High School Graduation Date

<i>Value</i>	<i>Unweighted Frequency</i>	<i>%</i>
BLANKED FOR CONFIDENTIALITY - ICPSR	2067	100.0 %

Based upon 2067 valid cases out of 2067 total cases.