August 7, 2003

ERRATA SHEET
National Household Education Surveys Program

There is an error in the derived variable LANGUAGE on the 2001 ASPA and ECPP data files (aspaasc.dat and ecppasc.dat respectively) and the 1999 Parent and Youth data files (99parent.dat and 99youth.dat, respectively).

This errata sheet lists the NHES products affected by the error, describes the error and its scope, and provides updated frequencies along with SAS, SPSS, and Stata code that can be used to correct the LANGUAGE variable on the data files.

NHES Products Affected

- NHES:2001 Data Files and Electronic Codebook CD-ROM (NCES 2003-078)
- NHES:2001 ASPA and ECPP Data Files Online (NCES 2003-045 and NCES 2003-046, respectively)
- NHES:2001 Data File User's Manual, Volume II, Early Childhood Program Participation Survey (NCES 2003-080)
- NHES:2001 Data File User's Manual, Volume III, Before- and After-School Programs and Activities Survey (NCES 2003-081)
- NHES:1991–1999 Data and Electronic Codebook CD-ROM (NCES 2002-005)
- NHES:1999 Parent and Youth Data Files Online (NCES 2000-105 and NCES 2000-106rv, respectively)
- NHES:1999 Data File User's Manual, Volume II — Parent Interview Data File (NCES 2000-081)
- NHES:1999 Data File User's Manual, Volume III — Youth Interview Data File (NCES 2000-082)
- Reports released prior to July 2003 that use the incorrectly derived variable LANGUAGE from the affected 1999 and 2001 datasets
  - *Trends in the Use of School Choice: 1993 to 1999* (NCES 2003-031)
  - *Condition of Education, 2003*, Indicator 38 (NCES 2003-067)

Description of the Error

The LANGUAGE variable is meant to indicate whether or not the parents/guardians in the household know English. A parent is categorized as knowing English if English was one of his/her first languages or one of the primary languages he/she spoke in the household at the time of the survey. The LANGUAGE variable included on the data files mistakenly categorized parents as knowing English if their first language was non-English and non-Spanish and their primary household language was their first language. For example, if a mother's first language was German (MOMLANG=91) and she

spoke mainly German in the household (MOMSPEAK=5), then she was incorrectly coded as knowing English.[1] Fathers were miscoded in the same way.

Scope of the Error

The 2001 ECPP has an unweighted sample size of 6,749. There are 154 cases that have a miscoded LANGUAGE variable:
- There are 28 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "one of two parents knows English" (LANGUAGE=2);
- There are 116 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "no parent knows English" (LANGUAGE=3); and
- There are 10 cases coded "one of two parents knows English" (LANGUAGE=2) that should have been coded "no parent knows English" (LANGUAGE=3).

The 2001 ASPA has an unweighted sample size of 9,583. There are 192 cases that have a miscoded LANGUAGE variable.
- There are 44 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "one of two parents knows English" (LANGUAGE=2).
- There are 134 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "no parent knows English" (LANGUAGE=3).
- There are 14 cases coded "one of two parents knows English" (LANGUAGE=2) that should have been coded "no parent knows English" (LANGUAGE=3).

The 1999 Parent survey has an unweighted sample size of 24,600. There are 561 cases that have a miscoded LANGUAGE variable:
- There are 111 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "one of two parents knows English" (LANGUAGE=2);
- There are 410 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "no parent knows English" (LANGUAGE=3); and
- There are 40 cases coded "one of two parents knows English" (LANGUAGE=2) that should have been coded "no parent knows English" (LANGUAGE=3).

---

[1] When MOMSPEAK=5, the mother's current primary household language(s) is/are the same as the first language(s) she learned. Thus, if MOMSPEAK=5 and MOMLANG=4 then the mother's current primary household languages are English and another (non-Spanish) language. If, however, MOMSPEAK=5 and MOMLANG=91, then the mother's primary household language is the other (non-English and non-Spanish) language she specified in MOMLANG. The language measures pertaining to fathers follow comparable logic.

The 1999 Youth survey has an unweighted sample size of 7,913. There are 181 cases that have a miscoded LANGUAGE variable:

- There are 32 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "one of two parents knows English" (LANGUAGE=2);
- There are 136 cases coded "all parents know English" (LANGUAGE=1) that should have been coded "no parent knows English" (LANGUAGE=3); and
- There are 13 cases coded "one of two parents knows English" (LANGUAGE=2) that should have been coded "no parent knows English" (LANGUAGE=3).

NHES-ASPA:2001 Published and Corrected Frequencies for LANGUAGE

NHES-ASPA:2001 Published Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	8,562	89.3	33,535,740	91.4
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	203	2.1	636,903	1.7
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	818	8.5	2,506,359	6.8
		9,583	**100.0%**	**36,679,001**	**100.0%**

NHES-ASPA:2001 Corrected Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	8,384	87.5	32,781,162	89.4
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	233	2.4	736,105	2.0
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	966	10.1	3,161,734	8.6
		9,583	**100.0%**	**36,679,001**	**100.0%**

NHES-ECPP:2001 Published and Corrected Frequencies for LANGUAGE

NHES-ECPP:2001 Published Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	5,861	86.8	18,142,381	89.5
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	175	2.6	399,450	2.0
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	713	10.6	1,739,394	8.6
		6,749	**100.0%**	**20,281,225**	**100.0%**

NHES-ECPP:2001 Corrected Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	5,717	84.7	17,698,377	87.3
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	193	2.9	442,448	2.2
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	839	12.4	2,140,400	10.6
		6,749	**100.0%**	**20,281,225**	**100.0%**

NHES-Parent:1999 Published and Corrected Frequencies for LANGUAGE

NHES-Parent:1999 Published Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	22,232	90.4	66,267,886	91.8
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	435	1.8	1,007,606	1.4
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	1,933	7.9	4,875,916	6.8
		24,600	**100.0%**	**72,151,408**	**100.0%**

NHES-Parent:1999 Corrected Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	21,711	88.3	64,837,647	89.9
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	506	2.1	1,195,216	1.7
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	2,383	9.7	6,118,545	8.5
		24,600	**100.0%**	**72,151,408**	**100.0%**

NHES-Youth:1999 Published and Corrected Frequencies for LANGUAGE

NHES-Youth:1999 Published Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	7,232	91.4	25,231,590	92.4
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	116	1.5	370,695	1.4
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	565	7.1	1,699,152	6.2
		7,913	**100.0%**	**27,301,437**	**100.0%**

NHES-Youth:1999 Corrected Frequencies for LANGUAGE

Response	Codes	Frequency	Percent	Weighted Frequency	Weighted Percent
1 BOTH SPEAK ENGLISH	1	7,064	89.3	24,622,777	90.2
2 ONE PRNT SPEAKS NON-ENGLISH LANG	2	135	1.7	445,564	1.6
3 BOTH PRNTS SPEAK NON-ENGLISH LANG	3	714	9.0	2,233,098	8.2
		7,913	**100.0%**	**27,301,439**	**100.0%**

Correction

Corrected frequencies for the LANGUAGE variable are available above. Below is the SAS, SPSS, and Stata code required to fix the variable in the 2001 ASPA and ECPP data files and the 1999 Parent and Youth data files.

To use the code to update the LANGUAGE variable on the 1999 Youth data file, it is necessary to link the Youth data file with the Parent file. This is because the variables MOMLANG, DADLANG, MOMSPEAK, and DADSPEAK were not included on the Youth data file. For more information on merging data files, please see Appendix G of the NHES:1999 Data File User's Manual, Volume III — Youth Interview Data File (NCES 2000-082), which is available at http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2000082.

SAS Code to Correct LANGUAGE

```
LANG_MOM=2;
if MOMLANG in (1,3,4) then LANG_MOM=1;
      else if MOMSPEAK in (1,3,4) then LANG_MOM=1;
      else if (MOMLANG=-1 & MOMSPEAK=-1) then LANG_MOM=-1;


LANG_DAD=2;
if DADLANG in (1,3,4) then LANG_DAD=1;
      else if DADSPEAK in (1,3,4) then LANG_DAD=1;
      else if (DADLANG=-1 & DADSPEAK=-1) then LANG_DAD=-1;


LANGUAGE=3;
if (LANG_MOM=1 & LANG_DAD=1) then LANGUAGE=1;
      else if (LANG_MOM=1 & LANG_DAD =-1) then LANGUAGE=1;
      else if (LANG_MOM=-1 & LANG_DAD=1) then LANGUAGE=1;
      else if (LANG_MOM=1 & LANG_DAD=2) then LANGUAGE=2;
      else if (LANG_MOM=2 & LANG_DAD=1) then LANGUAGE=2;
```


SPSS Code to Correct LANGUAGE

```
compute LANG_MOM=2.
do if (MOMLANG=1 or MOMLANG=3 or MOMLANG=4).
      compute LANG_MOM=1.
      else if (MOMSPEAK=1 or MOMSPEAK=3 or MOMSPEAK=4).
      compute LANG_MOM=1.
      else if (MOMLANG=-1 and MOMSPEAK=-1).
      compute LANG_MOM=-1.
      end if.

compute LANG_DAD=2.
do if (DADLANG=1 or DADLANG=3 or DADLANG=4).
      compute LANG_DAD=1.
      else if (DADSPEAK=1 OR DADSPEAK=3 OR DADSPEAK=4).
      compute LANG_DAD=1.
      else if (DADLANG=-1 AND DADSPEAK=-1).
      compute LANG_DAD=-1.
      end if.

Compute LANGUAGE=3.
do if (LANG_MOM=1 and LANG_DAD=1).
      compute LANGUAGE=1.
      else if (LANG_MOM=1 and LANG_DAD=-1).
      compute LANGUAGE=1.
      else if (LANG_MOM=-1 and LANG_DAD=1).
      compute LANGUAGE=1.
      else if (LANG_MOM=1 and LANG_DAD=2).
      compute LANGUAGE=2.
      else if (LANG_MOM=2 and LANG_DAD=1).
      compute LANGUAGE=2.
      end if.
```

STATA Code to Correct LANGUAGE

```
generate LANG_MOM=2;
replace LANG_MOM=1 if (MOMLANG==1 | MOMLANG==3 | MOMLANG==4);
replace LANG_MOM=1 if (MOMSPEAK==1 | MOMSPEAK==3 | MOMSPEAK==4);
replace LANG_MOM=-1 if (MOMLANG==-1 & MOMSPEAK==-1);

generate LANG_DAD=2;
replace LANG_DAD=1 if (DADLANG==1 | DADLANG==3 | DADLANG==4);
replace LANG_DAD=1 if (DADSPEAK==1 | DADSPEAK==3 | DADSPEAK==4);
replace LANG_DAD=-1 if (DADLANG==-1 & DADSPEAK==-1);

replace LANGUAGE=3;
replace LANGUAGE=1 if (LANG_MOM==1 & LANG_DAD==1);
replace LANGUAGE=1 if (LANG_MOM==1 & LANG_DAD==-1);
replace LANGUAGE=1 if (LANG_MOM==-1 & LANG_DAD==1);
replace LANGUAGE=2 if (LANG_MOM==1 & LANG_DAD==2);
replace LANGUAGE=2 if (LANG_MOM==2 & LANG_DAD==1);
```