Appendix 1: Sample SAS Syntax to Prepare the PIAAC Data for Mplus

```
/*********************************************************************;
************* PIAAC Data preparation for Mplus ************************;
*********************************************************************;
By "MASKED FOR BLIND REVIEW"
*/

*PIAAC = Program for International Assessment of Adult Competencies

*** Three stars indicate that the user input is required.

*Set the directory path;

libname in ''; *** Add the directory path between the single quotation marks.
               If you change the library name, reflect it in the rest of the program;
options nofmterr; *In case the format file is unavailable, run this option;

*Read the PIAAC Public Use File and create a temporary dataset;
*These are the examples for demonstration, users should refer to the PIAAC codebook,
and write new syntax as needed;
data temp;
   set in.prgushp1_puf;

*Motivation to learn construct items;
   MtL01 = I_Q04D;
   MtL02 = I_Q04J;
   MtL03 = I_Q04L;
   MtL04 = I_Q04M;

*Age group --> 5 year increments;
agecat12 = AGEG5LFSEXT;

*Gender --> 1 = female, 0 = male;
if GENDER_R = 1 then female = 0;
   if GENDER_R = 2 then female = 1;

*To change the reference group, use the following syntax (remove *);
*if GENDER_R = 1 then male = 1;
   *if GENDER_R = 2 then male = 1;

*Race --> A series of dichotomous race/ethnicity indicators;
if RACETHN_4CAT  = 2 then white = 1;
   if RACETHN_4CAT  in (1 3 6) then white = 0;

if RACETHN_4CAT  = 3 then black = 1;
   if RACETHN_4CAT  in (1 2 6) then black = 0;

if RACETHN_4CAT  = 1 then hisp = 1;
   if RACETHN_4CAT  in (2 3 6) then hisp = 0;

if RACETHN_4CAT  = 6 then others = 1;
   if RACETHN_4CAT  in (1 2 3) then others = 0;

*College education -> 1 = college or higher, 0 = less than college;
if B_Q01AUS_C = 3 then college = 1;
   if B_Q01AUS_C in (1 2) then college = 0;

*To change the reference group, use the following syntax (remove *);
*if B_Q01AUS_C = 2 then highschool = 1;
   *if B_Q01AUS_C in (1 3) then highschool= 0;

*if B_Q01AUS_C = 1 then lessthanHS = 1;
```

```
   *if B_Q01AUS_C in (2 3) then lessthanHS = 0;

*Reverse code self-rated health as --> 1-5 = poor - excellent;
if I_Q08_T = 1 then srh = 5;
   if I_Q08_T = 2 then srh = 4;
      if I_Q08_T = 3 then srh = 3;
         if I_Q08_T = 4 then srh = 2;
            if I_Q08_T = 5 then srh = 1;

*Create a subset of data containing only the variables of interest;
*Include all plausible values, sampling weights and replicate weights;
keep
   MtL01 MtL02 MtL03 MtL04 PVLIT1-PVLIT10
   agecat12 female white black hisp others college srh
   SPFWT0 SPFWT1-SPFWT80;
run;

*Check the coding results --> e.g., miminum, maximum, mean, etc. ;
proc means data = temp;
run;

*Limit the sample to adults aged 25 years and older;
*This example is for demonstration, users should refer to the codebook, and write new
syntax as needed;
data subbase;
   set temp;
   where agecat12 GE 3 ;
run;

*Replace all missing values with the missing value indicator = -9999;
data sub;
   set subbase;

*In Mplus, the missing values (. .N .D .R .V .M) are assigned values of -9999,
   which then will be assigned as missing in Mplus.;
array change _numeric_;
   do over change;
      if change = . or change in (.N .D .R .V .M)then change = -9999;
   end;
run;

*Create 10 datasets. each containing the variables of interest plus one set of
plausible values for PVLIT;
*The following macro program creates a dataset with one plausible value, and remove
all others.;

%MACRO plausible (origdata = , dataname = , dnum = );
data &dataname;
   set &origdata;

PVLIT = PVLIT&dnum;

drop
PVLIT1-PVLIT10

run;
%MEND plausible;

*Creating a macro variable for the directory where the .csv file will be saved;

%LET lib = YOUR_DIRECTORY; *** Replace YOUR_DIRECTORY to the user's folder locatin
(directory path);
*e.g., %LET lib = C:\Users\PIAAC ;
```

```
*Run the macro program --- from %MACRO to %MEND, and %LET --- first, then,
run %plausible with the specific parameters;
   *origdata is the temporary or permanent dataset name;
   *dataname is the name of new dataset;
   *dnum is the data number (e.g., plausible value 1 = data01);

*Dataset01;
%plausible (origdata = sub , dataname = Dataset01 , dnum = 1);

*Export the file as a csv file;
*Keep the variable names (PUTNAMES=YES) just for the dataset01 for the Mplus coding;

*Save one file with all variable names ("Dataset01WithNames.csv");
PROC EXPORT DATA= Dataset01
            OUTFILE= "&lib\Dataset01WithNames.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;

*The remaining .csv files to be used in Mplus do not include the variable names;
*Save one file without the variable names("Dataset01.csv");
PROC EXPORT DATA= Dataset01
            OUTFILE= "&lib\Dataset01.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Repeat the process (without the variable names) for the plausible values 2 - 10;

*Dataset02;
%plausible (origdata = sub , dataname = Dataset02 , dnum = 2);

*Save one file without the variable names ("Dataset02.csv");
PROC EXPORT DATA= Dataset02
            OUTFILE= "&lib\Dataset02.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset03;
%plausible (origdata = sub , dataname = Dataset03 , dnum = 3);

*Save one file without the variable names ("Dataset03.csv");
PROC EXPORT DATA= Dataset03
            OUTFILE= "&lib\Dataset03.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset04;
%plausible (origdata = sub , dataname = Dataset04 , dnum = 4);

*Save one file without the variable names ("Dataset04.csv");
PROC EXPORT DATA= Dataset04
            OUTFILE= "&lib\Dataset04.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset05;
%plausible (origdata = sub , dataname = Dataset05 , dnum = 5);
```

```
*Save one file without the variable names ("Dataset05.csv");
PROC EXPORT DATA= Dataset05
            OUTFILE= "&lib\Dataset05.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset06;
%plausible (origdata = sub , dataname = Dataset06 , dnum = 6);

*Save one file without the variable names ("Dataset06.csv");
PROC EXPORT DATA= Dataset06
            OUTFILE= "&lib\Dataset06.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset07;
%plausible (origdata = sub , dataname = Dataset07 , dnum = 7);

*Save one file without the variable names ("Dataset07.csv");
PROC EXPORT DATA= Dataset07
            OUTFILE= "&lib\Dataset07.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset08;
%plausible (origdata = sub , dataname = Dataset08 , dnum = 8);

*Save one file without the variable names ("Dataset08.csv");
PROC EXPORT DATA= Dataset08
            OUTFILE= "&lib\Dataset08.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset09;
%plausible (origdata = sub , dataname = Dataset09 , dnum = 9);

*Save one file without the variable names ("Dataset09.csv");
PROC EXPORT DATA= Dataset09
            OUTFILE= "&lib\Dataset09.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

*Dataset10;
%plausible (origdata = sub , dataname = Dataset10 , dnum = 10);

*Save one file without the variable names ("Dataset10.csv");
PROC EXPORT DATA= Dataset10
            OUTFILE= "&lib\Dataset10.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=NO;
RUN;

/*************************************************************;
************* End of SAS Syntax for Mplus Data preparation;
*************************************************************/
```