**ELFK Data entry R-Scripts Manual**

**Before using script**

1. Elvis shared drive -> space/ELFK/Data\_Entry/Questionnaires/R\_Cleaning\_Qualtrics\_Data
2. Download raw files from Qualtrics
   1. Questionnaire-> Data & Analysis -> Export & Import -> Export Data -> Download Data Table -> CSV -> Download all fields -> Use numerical values (most of the time) -> download
3. Save raw files in space/ELFK/Data\_Entry/Questionnaires/R\_Cleaning\_Qualtrics\_Data/ Qualtrics\_RAW\_FOR\_R
4. Make sure ID numbers do not include letters instead of numbers e.g. EL06**I**p should EL06**1**P

**Using script**

**Pay attention to green notes written in the scripts! Provide important info!**

1. Scripts already remove extra rows of fake/mock data BUT there may be additional mocks added since script was created.
2. When editing files to get rid of extra rows (new ones):
   1. Use script
   2. Look through the raw file and check for ID numbers such as “999”, “?”, “mock”, and other random entries… these are the extra rows you need to remove using: Questionnairename\_Q.# <- Questionnairename\_Q.# [ -c( row you want to remove), ]
   3. If unsure if an entry is a participant entry or mock data – check with Tricia
3. When editing files to get rid of extra columns:
   1. Use script
   2. First column is usually column 17/18 (ID number) e.g. Questionnairename\_Q.3 <- Questionnairename\_Q.2 [ , 17: last column needed]

**After script**

1. Always at the end of the script when cleaned excel file is created, open it and delete first column (rows not necessary)
2. Check for duplicate entries in the data. Edit script accordingly to remove/fix duplicates

***If any questionnaires are edited on Qualtrics and data is removed…. Merge both sheets to avoid having to manually input data from old sheet.***

***Convert NAs to blanks in cleaned sheets***

**Parent Questionnaire (PQ)**

Before performing Script

* PQ\_3 and PQ\_4
  + Check that data is in months, if not, change to months. Old data from latest extraction (**7/18/2018)** are already included in the script. Check new data only.

**Caregiving Questionnaire (CQ)**

\*\*\*\*EL106-EL108 data missing in Qualtrics

Data was deleted from Qualtrics when questionnaire was edited. **Do not erase** CQ raw file labeled as old data. This is the only record left of that data.

If new data is added after (7/18/18)- make sure to go down one at the very end of script where I add the same sex couple (notes in green in script) or else it’ll replace a new entry

**International Adoption Inventory (IAI)**

* IAI\_3\_AGEORPH
  + Manually add data entry codes to the script. Old data from latest extraction (**7/20/2018**) are already included in the script. Check new data only.
* IAI\_7\_AGEADOPT
  + If there are any entries in years instead of months, convert answers to months and add it to script. Already updated from last extraction on **7/20/2018.**
  + If any answers are in fractions (e.g. 10 ½ months) or in letters (e.g. ten months), you will have to edit these answers and add them to script. Already updated from last extraction on **7/20/2018**
* IAI\_12\_MONTHADJUST
  + Manually add data entry codes to the script. Old data from latest extraction (**7/5/2018**) are already included in the script. Check new data only.
* IAI\_19A\_OTHERINFO
  + This question was never on Qualtrics, so you have to manually enter this data based on data entry manual and data in IAI\_19B\_EXPLAIN. Already updated from last extraction on **720/2018**

**AIOI**

Before performing script!

* Check “Other” columns (Q2 and Q5) for valuable information that may change education choices. If “other” information requires change to education level, then change it. Once you assign a value in other education field, script will delete Q\_2 and Q\_5.

While performing script

* If they circle more than one education level for same parent- put the highest number only. Erase lower level education info. Just keep the entry in the highest degree option.
  + **Old data (since 7/20/2018 extraction) is already edited in script.**
  + **New data has to be reviewed**
* Any income that was inputted with k or m at the end e.g. 140k, will be read in r as 140, not 140,000. Remove any spaces in income too! Spaces for income messes up data during script! E.g. 150 000 should be rewritten as 150000.
* Check new data.
  + **Old data edits are in script (since 7/20/2018)**
  + **New data has to be reviewed**

After script

* Remove NAs and character(0) from clean file to blanks

**LEQ**

* LEQ\_MAX\_IMPACT
  + Look over new participants, find their highest impact score and enter this for last question LEQ\_MAX\_IMPACT
  + Updated with data from last extraction **6/28**/**2018**

**PPDS-F/PPDS-M**

1. Check questions Q10\_1 and Q10\_2:
   1. That respondent hasn’t already converted age so that we do not perform wrong calculations
      * 1. **Updated from extraction on 7/11/2017**
        2. Check the new data
   2. Convert any fractions into decimals or else script will convert fraction into 3 separate factors
2. Once you have run the script and created the cleaned excel file:
   1. PPDS\_F
      1. Replace all 0’s/NAs into blanks on output excel file for PPDS\_F\_7\_MENSTRUAGE – select “Find entire cells only” – or else it will delete all 0s e.g. 50 -> 5
      2. Replace all 0s into blanks on output excel file for PPDS\_F\_8\_HEIGHT - select “Find entire cells only”
      3. Replace all NAs into blanks in output excel sheet for PPDS\_F\_9\_WEIGHT
      4. Replace all NAs into blanks in output excel sheet for PPDS\_F\_DATE\_COMPLETE
   2. PPDS-M
      1. Replace all 0s into blanks on output excel file for PPDS\_M\_7\_HEIGHT - select “Find entire cells only”
      2. Replace all NAs into blanks in output excel sheet for PPDS\_M\_8\_WEIGHT
      3. Replace all NAs into blanks in output excel sheet for PPDS\_M\_DATE\_COMPLETE

**DSM**

* There are two versions (child version and PI parent version). Extract both from Qualtrics. Save separately.

**Demographics Questionnaire**

* If previous parents are asked to redo this questionnaire, make sure you delete the previous entry… which also means you will have to potentially fix script, section where you “FIX IDs that are incorrect” – check new data to make sure that there will be no new repeats/double entries. **Last updated on 7.13.2018**
* Replace all NAs into blanks on output excel file - select “Find entire cells only”