Assignment 2

For this assignment, a subset of 6 items (enj7: I enjoy math; good7: I am good at math; und7: I usually understand math; use7: math is useful in everyday problems; job7: math is needed for a good job; often7: I will use math often as an adult) for Grade 7 from the LSAY data set (N=2874) was used to conduct an exploratory factor analysis (EFA), exploratory factor analysis within a confirmatory factor analysis framework and a simple structure confirmatory factor analysis (CFA) to demonstrate that EFA within a CFA framework has the same maximum-likelihood chi-square value and degrees of freedom as an EFA. These six variables are measured on a 5-point Likert scale where 1-strongly agree, 2-agree, 3-not sure, 4-disagree, and 5 is strongly disagree.

Two factors were extracted in an EFA using the maximum likelihood method of estimation. Eigenvalues are greater than one for the first two components (2.857, 1.268). A factor with an eigenvalue greater than 1.0 indicates that the factor accounts for a greater amount of variance than had been contributed by one variable. This can be viewed graphically in the eigenvalue plot (Figure 1). The screeplot indicates a break between component 2 and 3. Since it starts to level off between the 2nd and 3rd factor, there is further evidence that there are only 2 components.

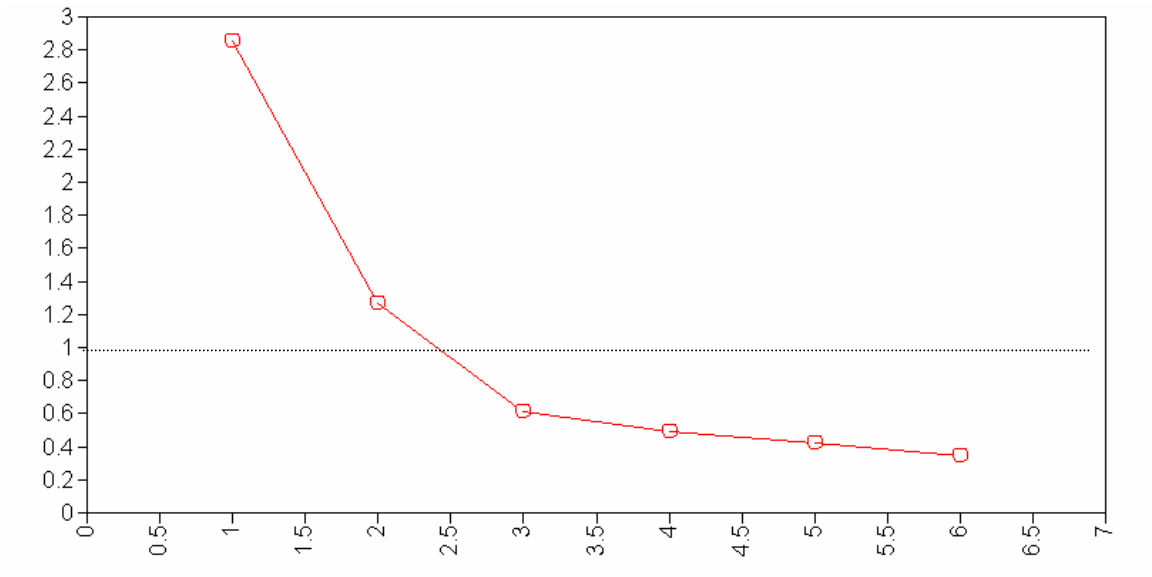

Figure 1. Eigenvalue plot for exploratory factor analysis.

Since the factors are correlated, a promax (oblique) rotation provides a conceptually clearer picture of the factors. Factor 1 includes items pertaining to affect or self efficacy toward mathematics (enj7, good7, und7). These three items load highly on Factor 1 (>.7) and do not load as highly on Factor 2 (<.08). These items have to do with the self confidence that one expresses in terms of math achievement or ability. Factor 2 includes items that pertain to utility or value of mathematics (use7, job7, often7). These three items load highly on Factor 2 (>.5) and do not load as highly on Factor 2 (<.06).

The EFA does not include standard errors to determine whether the factor loadings are statistically significant and does not include modification indices to determine if residual covariances are needed to represent minor factors[1]. These standard errors and modification indices can be obtained by carrying out an EFA in a CFA framework. Using the instructions provided in lecture, factor variances for the factor loadings of the two factors were set to 1 and factor loadings for two items were set to 0 for the two remaining restrictions. Good7 (an item that loaded highly on Factor 1 and low on Factor 2) and often7 (an item that loaded highly on Factor 2 and low on Factor 1) were the two anchor items selected to be set to 0.

The fit of this model is similar to the EFA and includes the same number of restrictions (Table 1). Table 1 indicates that the EFA and EFA within a CFA framework have the same chi-square values and degrees of freedom. The EFA within a CFA framework also provides additional fit indices which provide evidence that there is a good fit between the data and the model (i.e. CFI=0.999).

	Chi-Square			*Root mean square error of approximation*	*Root mean square residual*
	χ^2	*df*	*p*		
EFA	7.899	4	0.095	0.018	0.006
EFA within CFA framework	7.899	4	0.095	0.018	0.005
CFA with simple structure	72.416	8	0.000	0.053	0.029

Table 1. EFA (maximum likelihood estimation, promax rotation), EFA within a CFA framework and CFA with 2 factors for N=2874.

Factor determinacies from the EFA in a CFA framework indicate that Factor 1 has the bet determinacy because the factor loadings are higher than Factor 2 (0.915 versus 0.886). The variable regarding the use of math often as an adult (often7) has a large R^2 value indicating that it explains a large proportion of the observed variation. The standard errors for the factor loadings suggest that there is statistical significance. This is determined by dividing the estimate by the standard error. If this value is greater than 1.96, it suggests significant difference from 0. Of course, the two variables that were restricted to 0 are not significantly different from 0 (in Factor 1, often7 does not have a significant factor loading and in Factor 2, good7 does not have a significant factor loading).

[1] Joreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.
Muthen, B and Muthen, L. (2003). *Traditional latent variable modeling using Mplus: Mplus Short course notes*. California, Muthen & Muthen.
Van Prooijen, J., Van Der Kloot, W.A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement, 51*, 777-792.

The modification indices provide guidance on getting a better fit by correlating particular variables. The modification indices suggest that correlating job7 with enj7 (M.I.=6.917) and job7 with good7 (M.I.=5.089) might provide a better fit. This should be balanced with some substantive interpretation. For example if there is reason to believe that needing math for a good job is related to enjoying math, correlating these variables to get a better fit might be a good idea. In this case, believing that you are good at math might be linked to the belief that this skill is valued in the workforce. There are four modification indices that are statistically significant (job7 with enj7, job7 with good7, often7 with enj7, and often7 with use7) which suggests that those relationships should be freed and there is possibly a minor factor involved that allows these observed variables to correlate.

The CFA model does not fit better than the EFA or EFA in the CFA model. There are greater restrictions in the CFA and since we are using the same data, it is not surprising that there is a poor fit of this model. There is a higher chi-square, df and lower p-value in the CFA model as compared to the EFA model and EFA model in a CFA framework indicating that the CFA model does not fit the data. In future analyses, this should be an iterative process and information gained from one analyses should inform the other. But for the purpose of this assignment, a simple structure CFA with no modifications is what is reported.

As suggested, the same analysis was conducted separately for males and females. The rationale behind this is likely to examine whether the structure holds for males and females and will provide evidence to the idea that males and females relate differently to mathematics. Females and males might have different ideas about the value and use of mathematics and may report different levels of self efficacy. These analyses will help contribute to whether there are gender differences on these two factors. When conducting the same analysis separately by gender, results indicate that the 2 factor model fits better for males than for females. There are 1382 females and 1492 males included in this comparative analysis and a visual inspection of the means for each item seems to indicate roughly similar means. In addition to conducting separate analyses for males and females, as discussed in lab, a multiple group analysis which allows all parameters to be free was conducted and similar results were obtained. This analysis indicated that the model fits the data.

Looking at gender separately in the EFA, two factors were extracted. This was determined using similar criteria as the analyses that were not separated by gender (eigenvalues greater than 1 and examining screeplot). For males, the 2 factor model seems to fit better because the p-value is not significantly significant (p=0.2132) which indicates that the model fits the data well. This is not the necessarily the case for females (p=0.0924). Since chi-square values are sensitive to sample size (and there is a difference in sample size between males and females) and violations of the multivariate normality assumptions, examining the CFI and TLI also indicate that a small difference between males and females. It is not known whether this small difference is sufficient to consider the fit significantly different form males and females. A table (Table 2) similar to Table 1 is included after the MPlus input statements. The interpretation of the factors with promax rotation remains similar with enj7, good7, und7 loading highly on Factor 1 (and low loadings on Factor 2) and use7, job7, often7 loading highly on Factor 2 (and low loadings on Factor 1). The estimates are not terribly different for males and females which suggests a similar

structure of the data for males and females. The model fit for the EFA and EFA within a CFA framework are still similar. However, the slight difference in the fit between males and females might provide some evidence this structure fits the data from males better than the data from females. There may be two factors based on responses to these six items regarding self affect and efficacy toward math and the use and value of math in everyday life but this structure might be more appropriate or applicable to males than females.

MPlus Input Statements

Note: To explore differences between males and females, separate data files were created for males and females and then similar input statements were run. In addition, a multiple group analysis was conducted.

```
Title:  Week 2 Assignment: EFA

DATA: FILE IS lsay.dat;

VARIABLE:  NAMES ARE lsayid schcode classize urban tracking ntracks
              mthlvl female mthflg7-mthflg12 mothed fathed mothsei
              fathsei homeres race
              expect parapsh parcpsh parmpsh peerapsh peermpsh
              bas7 basse7 alg7 algse7 geo7 geose7
              qlt7 qltse7 mth7 mthse7 mtha7 mthase7
              bas8 basse8 alg8 algse8 geo8 geose8
              qlt8 qltse8 mth8 mthse8 mtha8 mthase8
              bas9 basse9 alg9 algse9 geo9 geose9
              qlt9 qltse9 mth9 mthse9 mtha9 mthase9
              bas10 basse10 alg10 algse10 geo10 geose10
              qlt10 qltse10 mth10 mthse10 mtha10 mthase10
              bas11 basse11 alg11 algse11 geo11 geose11
              qlt11 qltse11 mth11 mthse11 mtha11 mthase11
              bas12 basse12 alg12 algse12 geo12 geose12
              qlt12 qltse12 mth12 mthse12 mtha12 mthase12
              mthcrs7-mthcrs12 mtrk10-mtrk12 totstud lchfull
              lchpart parvis mcirr mclub strat mstrat comp mcomp
              african hispan asian expel arrest dropot self worth
              other satisf respect failure esteem problem cloctn
              dloctn eloctn floctn gloctn hloctn iloctn jloctn
              kloctn lloctn drink runawa suicid alc7 alc10 alc11
              alc12 arest7 runa8 runa9 runa10 runa11 run12 suic8
              suic9 suic10 suic11 suic12 drop7 drop8 drop9 drop10
              drop11 drop12 fdrop8 fdrop9 fdrop10 fdrop11 fdrop12
              enj7 good7 und7 useboy7 nerv7 wor7 scar7 use7 logic7
              boybet7 job7 often7 enj8 good8 und8 useboy8 nerv8
              wor8 scar8 use8 logic8 boybet8 job8 often8 enj9
              good9 und9 useboy9 nerv9 wor9 scar9 use9 logic9
              boybet9 job9 often9 enj10 good10 und10 useboy10
              nerv10 wor10 scar10 use10 logic10 boybet10 job10
              often10;
           USEVAR = enj7 good7 und7 use7 job7 often7;
           MISSING are all(9999);

  !VARIABLE DESCRIPTIONS:
  !enj7 = "I ENJOY MATH"
  !good7="I AM GOOD AT MATH"
  !und7= "USUALLY UNDERSTAND MATH"
  !use7 = "MATH USEFUL IN EVERYDAY PROBLEMS"
  !job7 ="NEED MATH FOR A GOOD JOB"
  !often7 = "WILL USE MATH OFTEN AS AN ADULT"
  !All of these variables have a 5-point Likert scale:
  !1-Strongly Agree, 2-Agree, 3-Not sure, 4-Disagree, 5-Strongly Disagree

ANALYSIS:
        type = EFA 1 3;
        estimator=ml;

OUTPUT:  sampstat;

PLOT:  Type is plot2;


- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Title:  Week 2 Assignment: CFA with Simple Structure.
```

```
DATA: FILE IS lsay.dat;

VARIABLE:  NAMES ARE lsayid schcode classize urban tracking ntracks
              mthlvl female mthflg7-mthflg12 mothed fathed mothsei
              fathsei homeres race
              expect parapsh parcpsh parmpsh peerapsh peermpsh
              bas7 basse7 alg7 algse7 geo7 geose7
              qlt7 qltse7 mth7 mthse7 mtha7 mthase7
              bas8 basse8 alg8 algse8 geo8 geose8
              qlt8 qltse8 mth8 mthse8 mtha8 mthase8
              bas9 basse9 alg9 algse9 geo9 geose9
              qlt9 qltse9 mth9 mthse9 mtha9 mthase9
              bas10 basse10 alg10 algse10 geo10 geose10
              qlt10 qltse10 mth10 mthse10 mtha10 mthase10
              bas11 basse11 alg11 algse11 geo11 geose11
              qlt11 qltse11 mth11 mthse11 mtha11 mthase11
              bas12 basse12 alg12 algse12 geo12 geose12
              qlt12 qltse12 mth12 mthse12 mtha12 mthase12
              mthcrs7-mthcrs12 mtrk10-mtrk12 totstud lchfull
              lchpart parvis mcirr mclub strat mstrat comp mcomp
              african hispan asian expel arrest dropot self worth
              other satisf respect failure esteem problem cloctn
              dloctn eloctn floctn gloctn hloctn iloctn jloctn
              kloctn lloctn drink runawa suicid alc7 alc10 alc11
              alc12 arest7 runa8 runa9 runa10 runa11 run12 suic8
              suic9 suic10 suic11 suic12 drop7 drop8 drop9 drop10
              drop11 drop12 fdrop8 fdrop9 fdrop10 fdrop11 fdrop12
              enj7 good7 und7 useboy7 nerv7 wor7 scar7 use7 logic7
              boybet7 job7 often7 enj8 good8 und8 useboy8 nerv8
              wor8 scar8 use8 logic8 boybet8 job8 often8 enj9
              good9 und9 useboy9 nerv9 wor9 scar9 use9 logic9
              boybet9 job9 often9 enj10 good10 und10 useboy10
              nerv10 wor10 scar10 use10 logic10 boybet10 job10
              often10;
           USEVAR = enj7 good7 und7 use7 job7 often7;
           MISSING are all(9999);


  !VARIABLE DESCRIPTIONS:
  !enj7 = "I ENJOY MATH"
  !good7="I AM GOOD AT MATH"
  !und7= "USUALLY UNDERSTAND MATH"
  !use7 = "MATH USEFUL IN EVERYDAY PROBLEMS"
  !job7 ="NEED MATH FOR A GOOD JOB"
  !often7 = "WILL USE MATH OFTEN AS AN ADULT"
  !All of these variables have a 5-point Likert scale:
  !1-Strongly Agree, 2-Agree, 3-Not sure, 4-Disagree, 5-Strongly Disagree

ANALYSIS: type=general;
          estimator = ml;

MODEL: f1 by enj7 good7 und7;
       f2 by use7 job7 often7;

OUTPUT: standardized modindices(0) sampstat FSDeterminacy;

PLOT:  Type is plot2;


- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -

Title:  Week 2 Assignment: EFA in a CFA framework

DATA: FILE IS lsay.dat;

VARIABLE:  NAMES ARE lsayid schcode classize urban tracking ntracks
              mthlvl female mthflg7-mthflg12 mothed fathed mothsei
              fathsei homeres race
              expect parapsh parcpsh parmpsh peerapsh peermpsh
              bas7 basse7 alg7 algse7 geo7 geose7
              qlt7 qltse7 mth7 mthse7 mtha7 mthase7
              bas8 basse8 alg8 algse8 geo8 geose8
```

```
              qlt8 qltse8 mth8 mthse8 mtha8 mthase8
              bas9 basse9 alg9 algse9 geo9 geose9
              qlt9 qltse9 mth9 mthse9 mtha9 mthase9
              bas10 basse10 alg10 algse10 geo10 geose10
              qlt10 qltse10 mth10 mthse10 mtha10 mthase10
              bas11 basse11 alg11 algse11 geo11 geose11
              qlt11 qltse11 mth11 mthse11 mtha11 mthase11
              bas12 basse12 alg12 algse12 geo12 geose12
              qlt12 qltse12 mth12 mthse12 mtha12 mthase12
              mthcrs7-mthcrs12 mtrk10-mtrk12 totstud lchfull
              lchpart parvis mcirr mclub strat mstrat comp mcomp
              african hispan asian expel arrest dropot self worth
              other satisf respect failure esteem problem cloctn
              dloctn eloctn floctn gloctn hloctn iloctn jloctn
              kloctn lloctn drink runawa suicid alc7 alc10 alc11
              alc12 arest7 runa8 runa9 runa10 runa11 run12 suic8
              suic9 suic10 suic11 suic12 drop7 drop8 drop9 drop10
              drop11 drop12 fdrop8 fdrop9 fdrop10 fdrop11 fdrop12
              enj7 good7 und7 useboy7 nerv7 wor7 scar7 use7 logic7
              boybet7 job7 often7 enj8 good8 und8 useboy8 nerv8
              wor8 scar8 use8 logic8 boybet8 job8 often8 enj9
              good9 und9 useboy9 nerv9 wor9 scar9 use9 logic9
              boybet9 job9 often9 enj10 good10 und10 useboy10
              nerv10 wor10 scar10 use10 logic10 boybet10 job10
              often10;
            USEVAR = enj7 good7 und7 use7 job7 often7;
            MISSING are all(9999);


  !VARIABLE DESCRIPTIONS:
  !enj7 = "I ENJOY MATH"
  !good7="I AM GOOD AT MATH"
  !und7= "USUALLY UNDERSTAND MATH"
  !use7 = "MATH USEFUL IN EVERYDAY PROBLEMS"
  !job7 ="NEED MATH FOR A GOOD JOB"
  !often7 = "WILL USE MATH OFTEN AS AN ADULT"
  !All of these variables have a 5-point Likert scale:
  !1-Strongly Agree, 2-Agree, 3-Not sure, 4-Disagree, 5-Strongly Disagree

ANALYSIS: estimator=ml;

MODEL: f1 by enj7-job7* often7@0;
       f2 by good7@0 enj7* und7* use7* job7* often7*;
       f1-f2@1;
       f2 with f1*;

OUTPUT: standardized modindices(0) sampstat fsdeterminacy;
```

	Chi-Square			Root mean square error of approximation	Root mean square residual	CFI	TLI	AIC	BIC
	χ^2	*df*	*p*						
EFA (Females)	7.968	4	0.0924	0.027	0.0097				
EFA (Males)	5.811	4	0.2132	0.017	0.0077				
EFA within CFA framework (Females)	7.968	4	0.0924	0.027	0.008	0.998	0.994	21177.338	21266.269
EFA within CFA framework (Males)	5.811	4	0.3243	0.017	0.006	0.999	0.998	22978.319	23068.553
CFA with simple structure for males and females	92.308	16	0.0000	0.058	0.032	0.986	0.973	44218.186	44373.236
CFA with simple structure (Females)	61.227	8	0.0000	0.069	0.038	0.978	0.958	21222.596	21290.603
CFA with simple structure (Males)	31.081	8	0.0001	0.044	0.025	0.992	0.985	22995.590	23064.592

Table 2. EFA (maximum likelihood estimation, promax rotation), EFA within a CFA framework and CFA with 2 factors for females (N=1382) and males (N=1492).