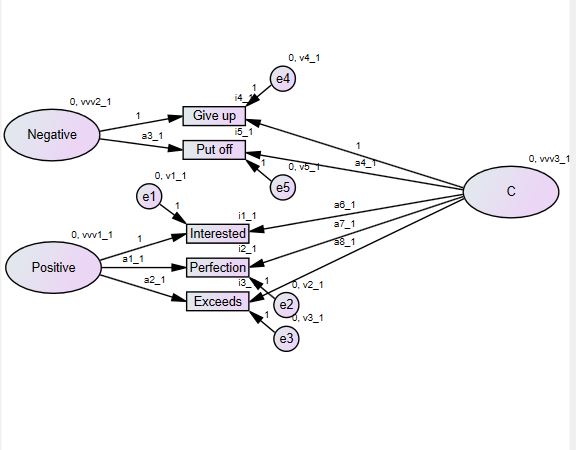
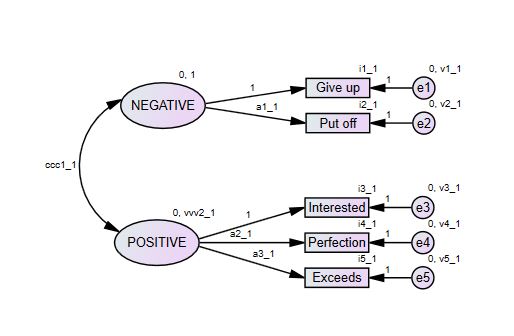
1. Extract data
2. I did some basic analysis, but Brent asked me to calculate the correlations between C and outcome variables (truancy and math performance), and SES
3. Brent asked me come up with a composite score for truancy (I took the mean; only 3 items)
4. There were some SES variables that can’t seem to be composited, so Brent said to use only parents’ education level, because “it seems to work the best”.
5. I ran a very quick ME analysis with the original data, by region (9 in total), but AMOS returned with an error message because of missing values – Brent asked me to first ditch all missing values, and then randomly sample 500 cases from each of the countries, and then run the ME analysis by region – a country got ditched as a whole because it has less than 500 cases.
6. First time running AMOS: a single factor model with C being the latent variable, and 5 items being the indicators – all models went well except for the unconstrained model, saying exceeding limit of iterations – won’t work even after I increase the limit for iterations.
7. I reported this to Brent, and he asked me to run a bi-factor model instead – C, positive & negative. First ran a quick EFA in SPSS, and C1 & C2 do load on Factor 1, while C3-C5 load on Factor 2 – amazing…
8. Then I tried to run bi-factor analysis in AMOS, but it keeps saying the model is unidentified and returning negative df – shoot! Then I realized I forgot to constrain one of the factor loadings to be 1 for C-to-indicators paths, and after I did that, the number of constraints needed dropped by 1, and the negative df changed from -9 to -8. Looks like this is what AMOS is talking about, but besides this factor loading, there does not seem to be a lot I can constrain… Reported this to Brent with a picture of the model used; hope he has a clue. Here’s the picture of the model:

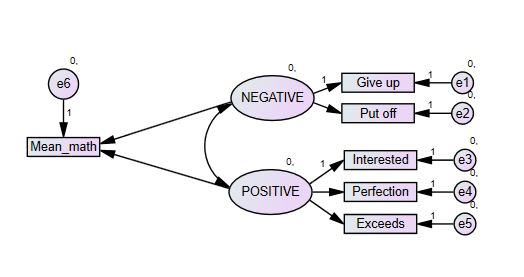
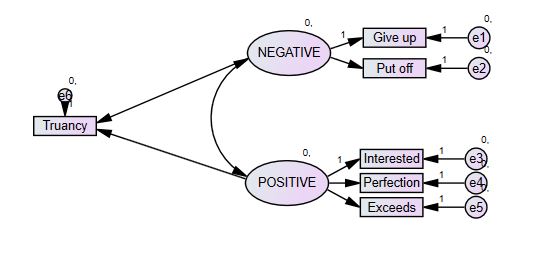
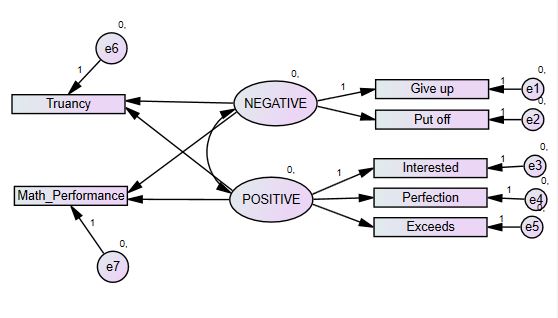


1. Brent suggested that I constrain some variances to make the model run and he suggested that I constrain all factor loadings and intercepts to run a fully constrained model – the models ran correctly except for the unconstrained model. Then Brent said he looked at the data and realized a clear 2-factor solution, and therefore I tried out a 2-factor solution – still need to put on one more constrain for the model to run besides default constraints. Here’s the model:



Finally, I saw OK in front of all models! CFIs indicate configural and metric equivalence, but not scalar, which is consistent with Brent’s guess. Problematic items are mostly Item 1 and Item 2 (negative items; give up and put off). I will probably have trouble trying to fix this, because they are the only 2 indicators that NEGATIVE have, and there’s no way I can free them both at the same time in AMOS, which requires that under each latent variable, at least one regression weight should be fixed to 1 (the “marked item”).

1. Then Brent asked me try a fully latent model (SEM model). And here’s the model:



Note the 1s (constraints). All models ran correctly (with the sweet “OK” in front of them).

1. Brent emailed to ask me to run the ME analyses across countries instead of regions. Unconstrained models are unidentified for both math and truancy, and therefore, for the truancy model, 5 more constraints on variances were placed (on the factor “Negative” for the first 5 groups), and 4 for the math model.

============================================================

**Updated on Feb 18, 2016**

1. So Brent asked if SES and gender were controlled for when I ran all these models, and I said no. So he asked me to rerun the analysis with with gender and SES controlled for. Besides this, he asked if I want to just controlling for gender, or run an analysis for female and male separately.
2. I felt uncomfortable fixing all error variance at 1, so I tried fixing two additional factor loadings to 1. However, Brent thought this was odd. He suggested I fix the latent variable variance to 1. I fixed the variance of Negative at 1, and I think it worked out, except for overall\_no\_USA\_math.

================================================================

**Updated Mar 20, 2016**

1. So Brent wants to do a meta-analysis with all the standardized regression weights from the measurement weights model, which he deemed as “the most cross-culturally comparable model”. He wants both the weights before and after the control for both dependent variables.
2. It’s kind of novel to do a meta-analysis with standardized regression weights instead of the correlations. There are some papers that I can refer to (saved) on this issue.

=========================================================================

**Updated on 09.13.2016:**

1. Brent asked me to compute the correlations between C and the predictors, both across countries and across the 9 major groups. It turns out the correlations changed direction when moving from the country level to the major group level (refer to 16.07.01 in the PISA private folder on PC).
2. It has been a long time since I updated this file last time, and a lot has happened.
3. I’ve moved to Mplus from AMOS, because Mplus is able to estimate latent means, and handle large samples. I was distracted by the simulation study with Fritz, and was unable to spend much time on the PISA project from April to June.
4. After I got back from China at the end of June, Brent emailed me saying he wants to restart the PISA project, and the first thing he wants me to do is to use Mplus. Mplus is actually a very well-developed SEM software, with big discussion forums on line, and tons of experts out there answering your questions. So I had little problem learning how to program my model under Mplus, and Eunike sent the code for estimating the means.
5. Eunike and I had some discussions over whether the errors have means, and she insisted that they don’t, but I think they do because they have variances. Anyways, the discussion ended with no agreement, but I did learn that you can’t estimate latent means and the intercepts at the same time, either (or some of either) should be fixed/constrained.
6. I ran the model where everything is constrained for truancy and math (i.e. an intercept model), respectively, and decided which parameters (weights and intercepts) to free across groups or for a few groups based on the modification indices. The truancy model worked out fine, and we have obtained some quite interesting results (refer to 09.12.2016 folder on Mac). However, the math model failed to compute positive covariance matrix, when I tried to free things based on the modification indices obtained from the all-constrained intercepts model with math as the predictor.
7. Brent suggested that
   1. Recode all the responses because the data used a reversed scale (1 = high, 5 = low), so that for the new items, 1 = low on this item, and 5 = high on this item (1 = don’t give up; 5 = give up easily) – which I’ve done already in 09.12.2016 folder on MAC.
   2. Run a measurement model only, and figure out which parameters are inequivalent across groups, and free those parameters for both the SEM with math and truancy as predictors. BY doing this, we’ve kept only the equivalent parts of the measurement constrained, and removed all the inequivalence. Therefore, the SEM we run after this will be more appropriate.
8. Then I started writing the paper (Jan 26 2017).