Multilevel Models Write-Up

Christopher Meng

12/16/2022

Part 1: Background

For the past few decades, U.S. education has been the site of a flurry of reforms in response to different political stimuli. With U.S.' status of "world superpower" under threat, politicians have turned to the U.S.' relative mediocrity on international test scores to identify education as one possible place of improvement to solidify its global standing. One well-established international assessment is the **Trends in International Mathematics and Science Study** (TIMSS).

While there are numerous issues with overemphasizing standardized testing as an evaluative assessment tool, the TIMSS dataset also collects substantive information on nationally representative samples of students, schools, and teachers from countries across the world. The breadth and depth of data collection provides an opportunity for people to make generalized comparisons between countries.

The first half of this semester, I reviewed probability distributions and different regression models and extended my understanding of likelihoods and other models and data structures. During the second half of the semester, I engaged with *hierarchical linear models* (HLM), which are also known as multilevel models or linear mixed effects models. HLMs are used to analyze data that is nested within each other. The term "levels" is often used to describe which level the data is collected, with smaller numbers indicating a more nested unit of analysis. For example, level 1-3 could be students, classrooms, and schools respectively. For a end-of-semester project, I decided to engage in a exploratory, comparative study between the U.S. and Taiwan using Grade 8 Math TIMSS 2015 data. My research questions were the following:

- Which factors most influence U.S. performance in Grade 8 Math TIMSS 2015 scores? Which factors most influence Taiwanese performance in Grade 8 Math TIMSS 2015 scores?
- What differences in students,teachers, and schools exist between Taiwan and the U.S.?

In order to answer these research questions, I attempted to fit two HLMs to the Grade 8 Math TIMSS 2015 data, one model for the U.S. and one model for Taiwan. Additionally, I created a Shiny app to summarize each variable in the dataset. In this write-up, I will talk about how I loaded in the data, summarize the dataset, present my model building process, and discuss the limitations, next steps, and reflections of the process.

Part 2: EDA

Data Setup

EdSurvey (Bailey et al. 2022) is a package created to read in and analyze data from the National Center for Education Statistics, which includes TIMSS data. The Grade 8 Math TIMSS 2015 data can also be directly downloaded from the website; this method will provide each individual dataset as well as codebooks, almanacs, and other additional information about the data. The code below reads in the desired data:

```r
TIMSS2015_US <- readTIMSS(path = "~/Downloads/TIMSS/2015", countries = c("usa"),
                          gradeLvl = "8", verbose = FALSE)
TIMSS2015_TW <- readTIMSS(path = "~/Downloads/TIMSS/2015", countries = c("twn"),
                          gradeLvl = "8", verbose = FALSE)
```

Using the Grade 8 codebook for 2015 ("TIMSS 2015 International Database" 2015), I selected variables, with their composite scale and subscale items, to explore in this project. I left out variables that did not fall under a composite scale, unless they were a descriptive, demographic variable. I identified the items using PDFs provided by TIMSS regarding the scales (linked below and cited in references). The following scales were selected, along with any demographic information collected (all scales have a range from 1-3; variables are the label in the codebook):

Scale_Name	Variable_Name	Level	Citations
Students Like Mathematics	BSDGSLM	1: Student	"Exhibit 10.4: Students Like Learning Mathematics" (2015)
Home Educational Resources	BSDGHER	1: Student	"Exhibit 4.2: Home Educational Resources" (2015)
Students' Sense of School Belonging	BSDGSSB	1: Student	"Exhibit 6.11: Students' Sense of School Belonging" (2015)
Engaging Teaching in Math Lessons	BSDGEML	1: Student	"Exhibit 10.2: Students' Views on Engaging Teaching in Mathematics Lessons" (2015)
Student Bullying	BSDGSB	1: Student	"Exhibit 7.6: Student Bullying" (2015)
Student Confidence in Mathematics	BSDGSCM	1: Student	"Exhibit 10.6: Students Confident in Mathematics" (2015)
Students Value Mathematics	BSDGSVM	1: Student	"Exhibit 10.7: Students Value Mathematics" (2015)
Weekly Time Spent on Math Homework	BSDMWKHW	1: Student	"Exhibit 9.8: Weekly Time Students Spend on Assigned Mathematics Homework" (2015)
School Emphasis on Academic Success	BTDGEAS	2: Teachers	"Exhibit 6.5: School Emphasis on Academic Success – Teachers' Reports" (n.d.)
Safe and Orderly Schools	BTDGSOS	2: Teachers	"Exhibit 7.4: Safe and Orderly School – Teachers' Reports" (2015)
School Conditions and Resources	BTDGSCR	2: Teachers	"Exhibit 5.9: Problems with School Conditions and Resources – Teachers' Reports" (2015)

Scale_Name	Variable_Name	Level	Citations
Teacher Job Satisfaction	BTDGTJS	2: Teachers	“Exhibit 6.7: Teacher Job Satisfaction” (2015)
Challenges Facing Teachers	BTDGCFT	2: Teachers	“Exhibit 6.9: Challenges Facing Teachers” (2015)
Teaching Limited by Student Needs	BTDGLSN	2: Teachers	“Exhibit 9.14: Teaching Limited by Student Needs” (2015)
Teachers Majored in Math or Math Ed	BTDM05	2: Teachers	NA
Instruction Affected by Math Resource Shortages	BCDGMRS	3: School	“Exhibit 5.7: Instruction Affected by Mathematics Resource Shortages – Principals’ Reports” (2015)
School Emphasis on Academic Success	BCDGEAS	3: School	“Exhibit 6.3: School Emphasis on Academic Success – Principals’ Reports” (2015)
School Discipline Problems	BCDGDAS	3: School	“Exhibit 7.2: School Discipline Problems – Principals’ Reports” (2015)

The level reflects who the questions were asked to rather than who the questions are talking about. For example, we see that there was two measures of school emphasis on academic success because the measure was given to both teachers and principals. Additionally, engaging math teaching is considered a level 1 variable because the measure was given to students. The following demographic variables were selected as well:

Demographics	Variables	Levels
Years of Teaching	BTBG01	2: Teachers
Sex of Teacher	BTBG02	2: Teachers
Age of Teacher	BTBG03	2: Teachers
Number of Students in the Class	BTBG12	2: Teachers
Number of Students with (Test) Language Difficulties	BTBG13	2: Teachers
School Composition by Student Economic Background	BCDG03	3: School
Percentage of Students Whose Native Language is the Language of the Test	BCBG04	3: School
Population Density/Urbanicity	BCBG05b	3: School
Free Breakfast	BCBG06a	3: School
Free Lunch	BCBG06b	3: School
Instructional Days per Year	BCBG07a	3: School
Instructional Minutes per Week	BCBG07b	3: School
Instructional Days per Week	BCBG07c	3: School
Place of Schoolwork Afterschool	BCBG08a	3: School
Student Achievement Used to Assign Math Classes	BCBG09a	3: School
Number of Computers	BCBG10	3: School
School Library	BCBG12	3: School
Difficulty Filling Math Teaching Vacancies	BCBG16a	3: School
Teacher Incentives for Retention	BCBG17a	3: School
Teacher Tardiness	BCBG18a	3: School
Teacher Absenteeism	BCBG18b	3: School
Years of Principal Experience	BCBG19	3: School

Demographics	Variables	Levels
Years as Principal at Current School	BCBG20	3: School

The `EdSurvey` package allows us to select a subset of variables. Additionally, it does the labor of merging the datasets. Initially, the downloaded data comes as separate files for teacher-, student-, and school-level variables for examples. `EdSurvey` addresses the merge process as well as the proper weighting of each observation. For brevity purposes, I omit the code for selecting the variables, since there are 212 variables. However, I used the `getData()` function from the `EdSurvey` package to save a data frame for each country. Using this function allows us to use the functions built into the package to summarize the data, as well as other tidy functions.

Data Overview

For the U.S., 246 schools ("Exhibit 5.13: School Sample Sizes – TIMSS 2015 – Eighth Grade" 2015) and 10,221 students ("Exhibit 5.15: Student Sample Sizes – TIMSS 2015 – Eighth Grade" 2015) were sampled. This analysis required each student to have a value for each variable considered; the final data frame after selecting the variables above has 4,458 observations.

For Taiwan, 190 schools ("Exhibit 5.13: School Sample Sizes – TIMSS 2015 – Eighth Grade" 2015) (including 100% retention from the original sample), and 5,711 students ("Exhibit 5.15: Student Sample Sizes – TIMSS 2015 – Eighth Grade" 2015) were sampled. The final data frame has 1,744 observations.

Teachers were surveyed on the basis that they taught this selected representative sample of students; thus, the TIMSS 2015 data is not a representative sample of 8th grade teachers. Schools were selected from the entire population of schools that had at least one 8th grade classroom. TIMSS 2015 employed a two-stage stratified sampling approach, selecting a representative sample of schools before selecting a representative sample of 8th graders. More technical details on the sampling process can be found here ("TIMSS - International Requirements for Sampling, Data Collection, and Response Rates" n.d.).

Rather than assess the entire set of testing items for every student, each student was given an intentional subset of the exam (i.e. missing data by design). However, each student completed too few items to assess a student individually, so the **plausible value** (PV) indicates the score that they would have had if they had taken the full exam. 5 PVs were randomly drawn from a distribution of students with similar background characteristics. When the PVs are aggregated, they provide unbiased estimates of the population score. However, again, conclusions should not be drawn about any single individual. Read more about PVs and how to use them here ("TIMSS - Weighting, Scaling, and Plausible Values" n.d.).

Univariate Summaries

Utilizing the Shiny application, I examined each potential variable and noted any observations of potential interest below. Most composite scales had 3 categories (1: lowest, 3: highest). I placed the category in brackets, e.g. [1], to clarify which level within the variable that I am discussing. (Unfortunately, due to limited memory storage on the free Shiny account, I do not have enough space to publish the application.)

The dependent variable of interest is the PVs for the Grade 8 Math TIMSS 2015 test scores. Taiwan has a noticeably higher mode PV around 650, while the U.S.' mode is nearly 100 points lower around 550. In fact, the distribution of Taiwan PVs has a slight ceiling effect, as evidenced by the skewed left distribution. This is repeated across the five PVs. Additionally, when we summarize the data by the categorical variable of which of 5 international benchmarks the PV score achieves (1 the lowest achievement and 5 the highest achievement), the plurality of 8th graders in Taiwan perform at the 5th international benchmark, while the plurality of 8th graders in the U.S. perform at the 3rd benchmark.

Notes about level 1 variables, surveying 8th grade math students:

- Noticeably more Taiwanese 8th graders (about 10% more) do not like learning mathematics [1] than U.S. 8th graders.
- About 15% more Taiwanese 8th graders report a sense of school belonging [2] than U.S. 8th graders, though about 10% more U.S. 8th graders scored a high sense of school belonging [3] than Taiwanese 8th graders.
- Almost 20% more Taiwanese 8th graders report “almost never” [1] being bullied.
- About 20% more U.S. 8th graders report “very engaging teaching” [3] than Taiwanese 8th graders, who reported around 15% more “engaging teaching” [2].
- Over 20% more Taiwanese 8th graders scored “not confident in mathematics” [1] than U.S. 8th graders.
- Almost 35% more U.S. 8th graders report “strongly valuing mathematics” [3] than Taiwanese 8th graders.
- On average, Taiwanese 8th graders say they do more hours of math homework per week than U.S. 8th graders.

Notes on level 2 variables, surveying 8th grade math teachers:

- There are more male teachers than female teachers in Taiwan, while the opposite holds for U.S. teachers.
- Taiwanese teachers tend to be middle-aged with 45% between the ages of 30-39, while a greater proportion of U.S. teachers than Taiwanese teachers are over 50.
- A greater proportion of Taiwanese teachers report having at least 1 student who has difficulties with the language of the test.
- 10% more Taiwanese teachers report a high emphasis on academic success in the school [2] compared to U.S. teachers. (A similar difference exists for principals at level 3.)
- Almost 10% more U.S. teachers report “less than safe and orderly” schools [1] than Taiwanese teachers.
- However, 10% more Taiwanese teachers report “minor problems” [2] with school conditions and resources than U.S. teachers.
- A little less than 10% more U.S. teachers were “less than satisfied” with their job [1] than Taiwanese teachers.
- U.S. teachers report having more challenges than Taiwanese teachers.
- A near-majority of Taiwanese teachers (just less than 50%) majored in only math, while U.S. teachers have a greater variety of majors.
- U.S. teachers reported more “very high” levels of confidence in teaching than Taiwanese teachers.

Notes on level 3 variables, surveying principals of the schools:

- The majority of Taiwanese schools (almost 70%) were classified as “neither affluent nor disadvantaged” [2], while over 50% of U.S. schools were considered “more disadvantaged” [1].
- Around 65% Taiwanese 8th graders’ native language was the language of the test, while just the same applied for just over 45% of U.S. 8th graders.
- Taiwanese schools do not typically provide free breakfast to students, whereas it is much more of a norm in U.S. schools.

- The mode of instructional days per school year is around 200 and 180 for Taiwanese and U.S. schools respectively. Both countries typically run 5 day school weeks, though U.S. schools appear to have, on average, slightly more instructional minutes per week than Taiwanese schools. *Note: Taiwan has a huge culture around "cram schools," where students as early as age 7 start attending intense afterschool to prepare for different standardized examinations. This may not be accounted for in this data, particularly regarding then number of instructional minutes per week.*
- While 70% of U.S schools use testing achievement to assign students to math classes, over 80% of Taiwanese schools do not use this practice.
- It appears to be more common that there are no math teaching vacancies to fill in Taiwan (almost 60%) compared to the U.S (over 40%). Additionally, U.S. principals report having a more difficult time filling vacant math teacher positions.
- Teacher tardiness and teacher absenteeism is reported as a nearly non-existent problem in Taiwanese schools, while it varied more across U.S. schools.
- The majority of Taiwanese schools report "hardly any problems" with discipline [1], while the majority of U.S. schools report "minor problems" with discipline [2].

Bivariate Summaries

I fit each predictor variable on the x-axis with the response variable of plausible value on the y-axis. The boxplots were a bit more difficult to discern noticeable differences in trends. However, there were a couple of interesting observations to note:

- The effect of a decreasing percentage of students whose native language is the language of the test is negative on the PV for Taiwan but displays virtually no association in the U.S. This suggests that, when there are few native speakers of the test language, Taiwanese 8th graders' PVs tended to be more negatively affected than 8th graders in the U.S.
- Similarly, urbanicity has no discernible association with PV for U.S. 8th graders. However, for Taiwanese 8th graders, those residing in a suburban area or a remote, rural area had noticeably lower PVs on average than students who resided in an urban area, a medium-sized city, or small town.

Part 3: Model

For HLMs, especially for models with greater than 2 levels, the number of parameters to estimate can increase exponentially with the inclusion of even one additional variable. Therefore, we usually have to articulate which random effects we are interested in modeling in the R code:

- (variable | group) indicates that we want to calculate variance for both the slope and the intercept and their covariance/correlation (i.e. random intercepts and random slopes). I believe this is the full specification.
- (1 | group) indicates that we only want the variance for the intercept (i.e. random intercepts with fixed mean model).
- (variable || group) OR (1 | group) + (0+variable | group) means that we want to calculate the variance for both the slope and the intercept but we assume their covariance/correlation is 0.

There are three strategies that researchers typically employ when modeling hierarchical data:

- Assume all error terms are independent (i.e. covariance/correlation is 0). This is the third one above.
- Begin with a random intercepts model, where only the first equation at each level (usually the intercept) has a error term where the variance needs to be estimated. This is the second one above.
- Reduce interaction terms.

For this project's purposes, I used a random intercepts model throughout my variable selection. In other words, I assumed that PV scores shared a common slope across schools and classrooms but allowed for random effects at the school and classroom levels in the intercept.

Unconditional Means Model

First, we fit a HLM for each country with no predictor variables. Below, we show the code for the first PV, but a separate model should be fit for each PV for a total of 10 models.

```
model0_US1 <- lmer(bsmmat01 ~ 1 + (1|idclass) + (1|idschool) , REML = TRUE, data = US_Math)
model0_TW1 <- lmer(bsmmat01 ~ 1 + (1|idclass) + (1|idschool) , REML = TRUE, data = TW_Math)
```

After fitting a separate model for each PV, we average the desired statistics (student variance, classroom variance, teacher variance, and the intercept which is the mean across all levels).

For Taiwan, the mean PV across all levels (students, classrooms, and schools) is 603.34. By dividing the estimated variance at each level by the total variance across levels, we can get an estimate of the total variability in PV explained at each level.

- 78.62 is the estimated variance in PV between Taiwanese 8th graders. 61.61% of the total variability in PV is explained at the student level in Taiwan.
- 3.70 is the estimated variance in PV between Taiwanese 8th grade classrooms. 2.90% of the total variability in PV is explained at the classroom level in Taiwan.
- 45.29 is the estimated variance in PV between Taiwanese schools. 35.49% of the total variability in PV is explained at the classroom level in Taiwan.

For the U.S., the mean PV across all levels (students, classrooms, and schools) is 519.16, over 80 points lower.

- 51.15 is the estimated variance in PV between U.S. 8th graders. 40.91% of the total variability in PV is explained at the student level in U.S.
- 58.10 is the estimated variance in PV between U.S. 8th grade classrooms. 46.47% of the total variability in PV is explained at the classroom level in U.S.
- 15.78 is the estimated variance in PV between U.S. schools. 12.62% of the total variability in PV is explained at the classroom level in U.S.

Already, we can see that the variability explained at the Taiwan classroom level is almost negligible, while the variability explained by variance in U.S. schools is much smaller than Taiwan schools.

Forward Selection

Because my analysis is exploratory, I employed a forward selection process to determine which variables to include in my model, starting from level 1 and building up to level 3. This is similar to the process suggested in this paper, which provides a nice overview of HLMs that covers much of what I discovered in this project (Anderson 2012). and makes intuitive sense as a starting point. The iterative process went as follows:

- Add each level 1 variable that is not currently selected to the previous model.
- Use `anova()` to compare the deviances between the previous model and each new model from the first step.
- Add the predictor variable that has the lowest p-value, assuming that there is a p-value under 0.05.
- Check for an interaction with the other predictors in the model by also seeing if the deviance drops significantly in `anova()`. Interactions were also examined if they were practically meaningful in interpretation; sometimes, because some variables have multiple categories beyond 3, the p-value appeared significant but possibly because the interaction added upwards of 12 additional terms, for example.
- Repeat the first four steps until there are no predictors that, when added to the model, return a significant drop in deviance.
- Repeat the process for levels 2 and 3.

Final US Model

After adding level 2 variables, the variance at level 3 (school) dropped close to 0. This indicates the the variance at the school level does not explain any additional variability in the overall model. Thus, no level 3 variables were chosen in the final model. While continuing variable selection indicated that there were more significant level 3 variables, they did not explain much variance at level 1 or level 2.

```
model_US <- lmer(bsmmat01 ~ bsdgscm + bsdgeml + bsdgher + bsdgsvm + bsdmwkhw + itsex +
                   bsdgscm:bsdgeml + btdgsos.math + btdglsn.math + btdgcft.math +
                   btbg03.math + btdgcft.math:bsdgher + (1|idclass) + (1|idschool),
                 REML = TRUE, data = US_Math)
summary(model_US)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: bsmmat01 ~ bsdgscm + bsdgeml + bsdgher + bsdgsvm + bsdmwkhw +
##     itsex + bsdgscm:bsdgeml + btdgsos.math + btdglsn.math + btdgcft.math +
##     btbg03.math + btdgcft.math:bsdgher + (1 | idclass) + (1 |      idschool)
##    Data: US_Math
##
## REML criterion at convergence: 47344.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3714 -0.6177  0.0256  0.6433  3.8755
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  idclass  (Intercept) 2.329e+03 4.826e+01
##  idschool (Intercept) 6.595e-07 8.121e-04
##  Residual             2.095e+03 4.577e+01
```

```
## Number of obs: 4458, groups:  idclass, 275; idschool, 137
##
## Fixed effects:
##                                                                       Estimate
## (Intercept)                                                           593.6273
## bsdgscmCONFIDENT IN MATHEMATICS                                       -31.0692
## bsdgscmNOT CONFIDENT IN MATHEMATICS                                   -61.3788
## bsdgemlENGAGING TEACHING                                                2.6285
## bsdgemlLESS THAN ENGAGING TEACHING                                     -4.7493
## bsdgherSOME RESOURCES                                                 -17.5324
## bsdgherFEW RESOURCES                                                  -28.6009
## bsdgsvmVALUE MATHEMATICS                                                0.4060
## bsdgsvmDO NOT VALUE MATHEMATICS                                        -8.7848
## bsdmwkhwMORE THAN 45 MINUTES BUT LESS THAN 3 HOURS                     -0.1385
## bsdmwkhw45 MINUTES OR LESS                                             -3.3251
## itsexMALE                                                               5.7422
## btdgsos.mathSAFE AND ORDERLY                                          -18.6335
## btdgsos.mathLESS THAN SAFE AND ORDERLY                                -24.3263
## btdglsn.mathSOMEWHAT LIMITED                                          -33.3942
## btdglsn.mathVERY LIMITED                                              -50.3112
## btdgcft.mathSOME CHALLENGES                                            -7.2043
## btdgcft.mathMANY CHALLENGES                                            34.0480
## btbg03.math25-29                                                       11.8309
## btbg03.math30-39                                                       -5.1387
## btbg03.math40-49                                                        7.4832
## btbg03.math50-59                                                       22.0162
## btbg03.math60 OR MORE                                                   7.1866
## bsdgscmCONFIDENT IN MATHEMATICS:bsdgemlENGAGING TEACHING                4.9850
## bsdgscmNOT CONFIDENT IN MATHEMATICS:bsdgemlENGAGING TEACHING            1.7030
## bsdgscmCONFIDENT IN MATHEMATICS:bsdgemlLESS THAN ENGAGING TEACHING      2.3289
## bsdgscmNOT CONFIDENT IN MATHEMATICS:bsdgemlLESS THAN ENGAGING TEACHING  10.7446
## bsdgherSOME RESOURCES:btdgcft.mathSOME CHALLENGES                      11.4494
## bsdgherFEW RESOURCES:btdgcft.mathSOME CHALLENGES                       18.8887
## bsdgherSOME RESOURCES:btdgcft.mathMANY CHALLENGES                       5.9717
## bsdgherFEW RESOURCES:btdgcft.mathMANY CHALLENGES                       13.1732
##                                                                       Std. Error
## (Intercept)                                                              14.4042
## bsdgscmCONFIDENT IN MATHEMATICS                                           2.4762
## bsdgscmNOT CONFIDENT IN MATHEMATICS                                       2.9443
## bsdgemlENGAGING TEACHING                                                  3.6214
## bsdgemlLESS THAN ENGAGING TEACHING                                        5.8745
## bsdgherSOME RESOURCES                                                     3.1580
## bsdgherFEW RESOURCES                                                      5.5428
## bsdgsvmVALUE MATHEMATICS                                                  1.5904
## bsdgsvmDO NOT VALUE MATHEMATICS                                           2.6794
## bsdmwkhwMORE THAN 45 MINUTES BUT LESS THAN 3 HOURS                        2.1036
## bsdmwkhw45 MINUTES OR LESS                                                2.3055
## itsexMALE                                                                 1.4318
## btdgsos.mathSAFE AND ORDERLY                                              5.9655
## btdgsos.mathLESS THAN SAFE AND ORDERLY                                   10.3634
## btdglsn.mathSOMEWHAT LIMITED                                              6.5519
## btdglsn.mathVERY LIMITED                                                 11.0817
## btdgcft.mathSOME CHALLENGES                                               6.7393
## btdgcft.mathMANY CHALLENGES                                              12.6161
```

```
## btbg03.math25-29                                                         15.6152
## btbg03.math30-39                                                         13.9145
## btbg03.math40-49                                                         14.1318
## btbg03.math50-59                                                         14.5230
## btbg03.math60 OR MORE                                                    15.9204
## bsdgscmCONFIDENT IN MATHEMATICS:bsdgemlENGAGING TEACHING                  4.3354
## bsdgscmNOT CONFIDENT IN MATHEMATICS:bsdgemlENGAGING TEACHING              4.5607
## bsdgscmCONFIDENT IN MATHEMATICS:bsdgemlLESS THAN ENGAGING TEACHING        6.6502
## bsdgscmNOT CONFIDENT IN MATHEMATICS:bsdgemlLESS THAN ENGAGING TEACHING    6.5637
## bsdgherSOME RESOURCES:btdgcft.mathSOME CHALLENGES                         4.0504
## bsdgherFEW RESOURCES:btdgcft.mathSOME CHALLENGES                          7.0852
## bsdgherSOME RESOURCES:btdgcft.mathMANY CHALLENGES                         6.0233
## bsdgherFEW RESOURCES:btdgcft.mathMANY CHALLENGES                         15.1384
##                                                                        t value
## (Intercept)                                                             41.212
## bsdgscmCONFIDENT IN MATHEMATICS                                        -12.547
## bsdgscmNOT CONFIDENT IN MATHEMATICS                                    -20.846
## bsdgemlENGAGING TEACHING                                                 0.726
## bsdgemlLESS THAN ENGAGING TEACHING                                      -0.808
## bsdgherSOME RESOURCES                                                   -5.552
## bsdgherFEW RESOURCES                                                    -5.160
## bsdgsvmVALUE MATHEMATICS                                                 0.255
## bsdgsvmDO NOT VALUE MATHEMATICS                                         -3.279
## bsdmwkhwMORE THAN 45 MINUTES BUT LESS THAN 3 HOURS                      -0.066
## bsdmwkhw45 MINUTES OR LESS                                              -1.442
## itsexMALE                                                                4.010
## btdgsos.mathSAFE AND ORDERLY                                            -3.124
## btdgsos.mathLESS THAN SAFE AND ORDERLY                                  -2.347
## btdglsn.mathSOMEWHAT LIMITED                                            -5.097
## btdglsn.mathVERY LIMITED                                                -4.540
## btdgcft.mathSOME CHALLENGES                                             -1.069
## btdgcft.mathMANY CHALLENGES                                              2.699
## btbg03.math25-29                                                         0.758
## btbg03.math30-39                                                        -0.369
## btbg03.math40-49                                                         0.530
## btbg03.math50-59                                                         1.516
## btbg03.math60 OR MORE                                                    0.451
## bsdgscmCONFIDENT IN MATHEMATICS:bsdgemlENGAGING TEACHING                 1.150
## bsdgscmNOT CONFIDENT IN MATHEMATICS:bsdgemlENGAGING TEACHING             0.373
## bsdgscmCONFIDENT IN MATHEMATICS:bsdgemlLESS THAN ENGAGING TEACHING       0.350
## bsdgscmNOT CONFIDENT IN MATHEMATICS:bsdgemlLESS THAN ENGAGING TEACHING   1.637
## bsdgherSOME RESOURCES:btdgcft.mathSOME CHALLENGES                        2.827
## bsdgherFEW RESOURCES:btdgcft.mathSOME CHALLENGES                         2.666
## bsdgherSOME RESOURCES:btdgcft.mathMANY CHALLENGES                        0.991
## bsdgherFEW RESOURCES:btdgcft.mathMANY CHALLENGES                         0.870
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

U.S.' and Taiwan's models both include the following level 1 variables: Student Confidence, Home Educational Resources, Value Mathematics, and Amount of Weekly Homework.

Engaging Math Teaching and Sex are unique variables to the U.S. Student Like Mathematics is the only unique variable to Taiwan.

U.S.' model accounted for level 2 but not level 3 variables, while Taiwan's model accounted for level 3 variables but not level 2.

Level	Variance_Unconditional	Variance_AfterStudent	Variance_FullModel
student	50.73	45.79	45.7700
classroom	57.65	51.97	48.2600
school	16.28	11.52	0.0008

Above are the student-, classroom-, and school-level variances of the unconditional means model, the model after adding student variables, and the final model.

A pseudo-R^2 value can be calculated by finding the difference between the variance at the level of interest and dividing by the variance of the unconditional means model. 9.73%, 9.84%, and 29.24% of additional variability in PV is explained at the student, classroom, and school level respectively after adding student-level predictors to the model.

After adding teacher-level variables, an additional 7.14% of variability in PV was explained at the classroom level, and nearly 100% of the variability in PV was explained at the school level.

Final Taiwan Model

From the unconditional means model, the variance at level 2 was close to 0. Thus, no variables were selected at that level.

```
model_TW <- lmer(bsmmat01 ~ bsdmwkhw + bsdgslm + bsdgscm + bsdgsvm + bsdgher +
                   bsdgslm:bsdgher + bcdgeas + bcbg04 + bcdg03 + (1|idclass) + (1|idschool),
                 REML = TRUE, data = TW_Math)
summary(model_TW)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: bsmmat01 ~ bsdmwkhw + bsdgslm + bsdgscm + bsdgsvm + bsdgher +
##     bsdgslm:bsdgher + bcdgeas + bcbg04 + bcdg03 + (1 | idclass) +
##     (1 | idschool)
##    Data: TW_Math
##
## REML criterion at convergence: 19452.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.5721 -0.5595  0.0756  0.6657  3.1494
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  idclass  (Intercept) 2.375e-04  0.01541
##  idschool (Intercept) 7.366e+02 27.13950
##  Residual             4.211e+03 64.88844
## Number of obs: 1744, groups:  idclass, 62; idschool, 61
##
## Fixed effects:
##                                                         Estimate
## (Intercept)                                              757.519
## bsdmwkhwMORE THAN 45 MINUTES BUT LESS THAN 3 HOURS         8.144
## bsdmwkhw45 MINUTES OR LESS                                -1.590
```

```
## bsdgslmLIKE LEARNING MATHEMATICS                                  14.360
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS                            4.904
## bsdgscmCONFIDENT IN MATHEMATICS                                  -26.484
## bsdgscmNOT CONFIDENT IN MATHEMATICS                              -79.096
## bsdgsvmVALUE MATHEMATICS                                           6.386
## bsdgsvmDO NOT VALUE MATHEMATICS                                  -22.638
## bsdgherSOME RESOURCES                                             -9.714
## bsdgherFEW RESOURCES                                             -19.556
## bcdgeasHIGH EMPHASIS                                             -49.332
## bcdgeasMEDIUM EMPHASIS                                           -61.859
## bcbg0476 TO 90%                                                  -13.048
## bcbg0451 TO 75%                                                    1.117
## bcbg0426 TO 50%                                                  -25.317
## bcbg0425% OR LESS                                                -46.855
## bcdg03NEITHER MORE AFFLUENT NOR MORE DISADVANTAGED               -23.614
## bcdg03MORE DISADVANTAGED                                         -36.521
## bsdgslmLIKE LEARNING MATHEMATICS:bsdgherSOME RESOURCES           -15.054
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS:bsdgherSOME RESOURCES  -16.332
## bsdgslmLIKE LEARNING MATHEMATICS:bsdgherFEW RESOURCES            -18.265
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS:bsdgherFEW RESOURCES   -51.656
##                                                                Std. Error
## (Intercept)                                                        20.189
## bsdmwkhwMORE THAN 45 MINUTES BUT LESS THAN 3 HOURS                  4.830
## bsdmwkhw45 MINUTES OR LESS                                          5.307
## bsdgslmLIKE LEARNING MATHEMATICS                                   12.077
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS                            12.332
## bsdgscmCONFIDENT IN MATHEMATICS                                     6.513
## bsdgscmNOT CONFIDENT IN MATHEMATICS                                 7.059
## bsdgsvmVALUE MATHEMATICS                                            5.790
## bsdgsvmDO NOT VALUE MATHEMATICS                                     6.366
## bsdgherSOME RESOURCES                                              11.457
## bsdgherFEW RESOURCES                                               19.447
## bcdgeasHIGH EMPHASIS                                               16.268
## bcdgeasMEDIUM EMPHASIS                                             16.617
## bcbg0476 TO 90%                                                    11.067
## bcbg0451 TO 75%                                                    13.716
## bcbg0426 TO 50%                                                    19.744
## bcbg0425% OR LESS                                                  16.288
## bcdg03NEITHER MORE AFFLUENT NOR MORE DISADVANTAGED                 11.520
## bcdg03MORE DISADVANTAGED                                           16.435
## bsdgslmLIKE LEARNING MATHEMATICS:bsdgherSOME RESOURCES             13.377
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS:bsdgherSOME RESOURCES      13.101
## bsdgslmLIKE LEARNING MATHEMATICS:bsdgherFEW RESOURCES              21.996
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS:bsdgherFEW RESOURCES       20.982
##                                                                t value
## (Intercept)                                                     37.521
## bsdmwkhwMORE THAN 45 MINUTES BUT LESS THAN 3 HOURS               1.686
## bsdmwkhw45 MINUTES OR LESS                                      -0.300
## bsdgslmLIKE LEARNING MATHEMATICS                                 1.189
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS                          0.398
## bsdgscmCONFIDENT IN MATHEMATICS                                 -4.067
## bsdgscmNOT CONFIDENT IN MATHEMATICS                            -11.205
## bsdgsvmVALUE MATHEMATICS                                         1.103
## bsdgsvmDO NOT VALUE MATHEMATICS                                 -3.556
```

```
## bsdgherSOME RESOURCES                                              -0.848
## bsdgherFEW RESOURCES                                               -1.006
## bcdgeasHIGH EMPHASIS                                               -3.033
## bcdgeasMEDIUM EMPHASIS                                             -3.723
## bcbg0476 TO 90%                                                    -1.179
## bcbg0451 TO 75%                                                     0.081
## bcbg0426 TO 50%                                                    -1.282
## bcbg0425% OR LESS                                                  -2.877
## bcdg03NEITHER MORE AFFLUENT NOR MORE DISADVANTAGED                 -2.050
## bcdg03MORE DISADVANTAGED                                           -2.222
## bsdgslmLIKE LEARNING MATHEMATICS:bsdgherSOME RESOURCES             -1.125
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS:bsdgherSOME RESOURCES  -1.247
## bsdgslmLIKE LEARNING MATHEMATICS:bsdgherFEW RESOURCES              -0.830
## bsdgslmDO NOT LIKE LEARNING MATHEMATICS:bsdgherFEW RESOURCES   -2.462
```

The equations would be far too long to write out. However, I will note that the level 2 equations would simply be $a_{ij} = a_i$ for each slope value needed in the level 1 equation, *except* for the first equation for the level 1 intercept which would also specify a variance. Similarly, only the first level 3 equation, $a_i = \alpha_0 + ...$ would have an additional error term distributed with a particular variance. The variances parameter estimations are listed above under "random effects." In addition looking at the additional variability explained by each level of variables below, I will interpret a couple values. For students who do not value mathematics (the lowest score on the composite scale), the change in PV test score for students, on average, is -22.64. A lack of confidence in mathematics is also associated with, on average, a 79.10 drop in PV test score for students. There is also a strong interaction between students not liking math learning and having few resources; for students who fall into both categories, the effect of one variable or the other on the PV test score for students was, on average, a drop of 51.66 points.

Level	Variance_Unconditional	Variance_AfterStudent	Variance_FullModel
student	78.07	64.89	64.89
classroom	0.00	0.00	0.00
school	46.33	36.27	27.14

Above are the student, classroom, and school level variances of the unconditional means model, the model after adding the student-level variables, and the final model. The variance at the classroom-level is arbitrarily close to 0 for all models.

After adding student-level variables, 16.88% and 21.72% of variability at the student and school levels respectively was accounted for.

After adding school-level variables, an additional 25.17% of variability at the school-level was explained.

Part 4: Discussion

By fitting the hierarchical models, the most significant observation was that variance at the classroom level in Taiwan did not significantly contribute to explaining variability in PVs (even initially starting very close to 0 without any predictors), while variance at the school level in the U.S. quickly converged to 0 as we took into account level 1 and level 2 variables. As a reminder, 0 variance in a HLM does not imply that there is no variance at the school level; clearly, each school scores somewhat differently on a standardized exam. However, it does say that the amount of school variance can be explained by accounting for variables at other levels in this case; essentially, one possible interpretation is that the hierarchy does not need to include the school level in this case. On the other hand, considering how Taiwan does not seem to have trouble recruiting math teachers and filling their positions, the U.S. well-documented struggles with finding qualified math teachers potentially corroborates the high amount of classroom-to-classroom variance.

I'm curious if this points to how classrooms in Taiwan are potentially more standardized, and the variability in test performance is attributed more to the diversity of the student population and, to a lesser degree, the variance across schools. Taiwan has a surprising (to a foreigner who does not know much) amount of languages and ethnicities, so the student-to-student variance adds an intriguing layer to the importance of considering those different backgrounds in Taiwan.

Limitations

There are many (many) limitations to this project. First of all, Anderson (2012) points to several assumptions, like other forms of regression, that one should check to determine the interpretability of the HLM. Beyond checking and addressing assumptions, ideally, I would have fit a final model to each of the 5 plausible values and averaged the parameter estimates. I also was wondering how other researchers dealt with model building, as I could only select my variables using the first PV as my response variable. Finally, there are limitations to using forward selection to pick the predictors in the final model. According to Shalizi (n.d.), "variable selection using p-values is a bad idea" because of numerous reasons that could render the p-values misleading. A significant variable might not have a significant p-value; an insignificant variable might have an inflated p-value. They suggest using cross-validation, if our interests lie in prediction, as a better method for determining which predictors to choose. Forward selection also has a blind spot for leaving out variables that may be correlated with each other.

Next Steps

Due to the global, longitudinal nature of this large dataset, there are many possibilities for further analysis. For example, I only analyzed 2015 data, though the data has been collected for a couple of decades every 4 years now. Additionally, it may be interesting to break down the plausible testing score into subscores in each domain reflecting different mathematical scores, which TIMSS data does collect. Using generalized linear mixed models, we also might have considered international benchmark achieved as a potential categorical response, rather than being limited to the continuous response variable of plausible test score.

Beyond the possibilities of the dataset, it could also be interesting to explore the psychometrics of looking at the entire scale (which is what I chose to do here) vs. individual items. How reliable would it be to use a single scale item as a predictor variable? Furthermore, some items appeared that they could belong together, but no scale was formally developed by TIMSS to calculate a composite index for those items. I left those variables out of this analysis, but constructs such as teacher confidence and teacher pedagogy could be interesting to examine.

In this project, I also would have liked to do the following with more time and energy. While Taiwan and the U.S. had only 2% and 4% of students fall under this category respectively, there were a not insignificant amount of students whose mathematics achievement was achievement was too low to estimate ("Appendix D.2: Percentages of Students with Achievement Too Low for Estimation*" 2015). Were these students dropped from the dataset? Are there other potentially meaningful patterns in the missing data of the existing dataset?

Finally, I would like to continue developing my understanding of the mathematical theory behind HLM. Developing content knowledge and reading further papers on HLM could help me figuring out which variance parameters might be meaningful to include beyond the random intercepts. In addition, I want to develop more tools to troubleshoot common "errors/warnings" that I ran into (e.g. failed to converge, singular). I understand singular to mean that either the variance is estimated to be 0 or the correlation is at the boundary of 1 or -1. The first is not necessarily a bad thing that needs to be adjusted, but it could signal overfitting, for example. I did significantly cut down on the number of variables in my final model, but we can still see that the number of estimated parameters is still massive, even after specifying the model as a random intercepts model.

References

Anderson, Daniel. 2012. "Hierarchical Linear Modeling (HLM): An Introduction to Key Concepts Within Cross-Sectional and Growth Modeling Frameworks." Technical {Report} 1308. University of Oregon: Behavioral Research; Teaching. https://files.eric.ed.gov/fulltext/ED545279.pdf.

"Appendix D.2: Percentages of Students with Achievement Too Low for Estimation*." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/11.-appendices/D_2_math-percentages-of-students-with-achievement-too-low-for-estimation-grade-8.pdf.

Bailey, Paul, Ahmad Emad, Huade Huo, Michael Lee, Yuqi Liao, Alex Lishinski, Trang Nguyen, et al. 2022. "EdSurvey: Analysis of NCES Education Survey and Assessment Data." https://CRAN.R-project.org/package=EdSurvey.

"Exhibit 10.2: Students' Views on Engaging Teaching in Mathematics Lessons." 2015. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/10.-student-engagement-and-attitudes/10_2_math-students-views-on-engaging-teaching-grade-8.pdf.

"Exhibit 10.4: Students Like Learning Mathematics." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/10.-student-engagement-and-attitudes/10_4_math-students-like-learning-mathematics-grade-8.pdf.

"Exhibit 10.6: Students Confident in Mathematics." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/10.-student-engagement-and-attitudes/10_6_math-students-confident-in-mathematics-grade-8.pdf.

"Exhibit 10.7: Students Value Mathematics." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/10.-student-engagement-and-attitudes/10_7_math-students-value-mathematics-grade-8.pdf.

"Exhibit 4.2: Home Educational Resources." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/4.-home-environment-support/4_2_math-home-educational-resources-grade-8.pdf.

"Exhibit 5.13: School Sample Sizes – TIMSS 2015 – Eighth Grade." 2015. https://timssandpirls.bc.edu/publications/timss/2015-methods/pdf/T15_MP_Exh5-13_School_Sample_Sizes_G8.pdf.

"Exhibit 5.15: Student Sample Sizes – TIMSS 2015 – Eighth Grade." 2015. TIMSS. https://timssandpirls.bc.edu/publications/timss/2015-methods/pdf/T15_MP_Exh5-15_Student_Sample_Sizes_G8.pdf.

"Exhibit 5.7: Instruction Affected by Mathematics Resource Shortages – Principals' Reports." 2015. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/5.-school-composition-and-resources/5_7_math-instruction-affected-by-mathematics-resource-shortages-grade-8.pdf.

"Exhibit 5.9: Problems with School Conditions and Resources – Teachers' Reports." 2015. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/science/5.-school-composition-and-resources/5_9_science-problems-with-school-conditions-and-resources-grade-8.pdf.

"Exhibit 6.11: Students' Sense of School Belonging." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/6.-school-climate/6_11_math-students-sense-of-school-belonging-grade-8.pdf.

"Exhibit 6.3: School Emphasis on Academic Success – Principals' Reports." 2015. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/6.-school-climate/6_3_math-school-emphasis-on-academic-success-principals-reports-grade-8.pdf.

"Exhibit 6.5: School Emphasis on Academic Success – Teachers' Reports." n.d. TIMSS. Accessed December 26, 2022. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/6.-school-climate/6_5_math-school-emphasis-on-academic-success-teachers-reports-grade-8.pdf.

"Exhibit 6.7: Teacher Job Satisfaction." 2015. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/science/6.-school-climate/6_7_science-teacher-job-satisfaction-grade-8.pdf.

"Exhibit 6.9: Challenges Facing Teachers." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/6.-school-climate/6_9_math-challenges-facing-teachers-grade-8.pdf.

"Exhibit 7.2: School Discipline Problems – Principals' Reports." 2015. TIMSS. https://timssandpirls.

bc.edu/timss2015/international-results/wp-content/uploads/filebase/science/7.-school-safety/7_2_science-school-discipline-problems-grade-8.pdf.

"Exhibit 7.4: Safe and Orderly School – Teachers' Reports." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/science/7.-school-safety/7_4_science-safe-and-orderly-school-grade-8.pdf.

"Exhibit 7.6: Student Bullying." 2015. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/7.-school-safety/7_6_math-student-bullying-grade-8.pdf.

"Exhibit 9.14: Teaching Limited by Student Needs." 2015. TIMSS. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/science/9.-classroom-instruction/9_14_science-teaching-limited-by-student-needs-grade-8.pdf.

"Exhibit 9.8: Weekly Time Students Spend on Assigned Mathematics Homework." 2015. https://timssandpirls.bc.edu/timss2015/international-results/wp-content/uploads/filebase/mathematics/9.-classroom-instruction/9_8_math-time-students-spend-on-mathematics-homework-grade-8.pdf.

Shalizi, Cosma. n.d. "Lecture 26: Variable Selection." Carnegie Mellon University. Accessed December 26, 2022. https://www.stat.cmu.edu/~cshalizi/mreg/15/lectures/26/lecture-26.pdf.

"TIMSS - International Requirements for Sampling, Data Collection, and Response Rates." n.d. *National Center for Education Statistics*. Accessed December 26, 2022. https://nces.ed.gov/timss/timss15technotes_intlreqs.asp.

"TIMSS - Weighting, Scaling, and Plausible Values." n.d. Accessed December 26, 2022. https://nces.ed.gov/timss/timss15technotes_weighting.asp.

"TIMSS 2015 International Database." 2015. *TIMSS & PIRLS International Study Center*. https://timssandpirls.bc.edu/timss2015/index.html.