Assignment 5

For this assignment, 12 items measuring math attitudes in Grade 7 (enj7: I enjoy math; good7: I am good at math; und7: I usually understand math; useboy7: Math is more useful for boys; nerv7: Math makes me nervous; wor7: I worry about math test grades; scar7: scared when I open a math book; use7: Math is useful in everyday problems; logic7: Math helps logical thinking; boybet7: Boys are better at math than girls; job7: Math is needed for a good job; often7: I will use math often as an adult) from the LSAY data set (N=3116) were used to conduct an exploratory factor analysis (EFA) and latent class analysis (LCA) in Mplus Version 3.01. Of the 3116 cases, 49 cases were missing values on all 12 items and were excluded. There were 3067 cases used in these analyses and missing values were assumed to be missing at random.

These 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. I dichotomized all variables with a 3, 4 or 5 (Not sure, Disagree, Strongly Disagree) considered a 0 (Disagree) and a 1 or 2 (Agree, Strongly Agree) as a 1 (Agree). I considered a 3-Likert response of "Not Sure" to mean that the respondent disagreed. The 0 category are for respondents who endorse the item by agreeing or strongly agreeing. The 1 categories are for respondents who do not endorse the item or are not sure. Table 1 provides the frequencies of the 12 items.

	Disagree//Not Sure	*Agree*
enj: I enjoy math	960 (31.4%)	2100 (68.6%)
good: I am good at math	931 (30.5%)	2126 (68.6%)
und: I usually understand math	700 (23.2%)	2320 (76.8%)
useboy: Math is more useful for boys	2636 (87.3%)	383 (12.7%)
nerv: Math makes me nervous	2261 (76.4%)	698 (23.6%)
wor: I worry about math test grades	1012 (33.8%)	1980 (66.2%)
scar: scared when I open a math book	2397 (78.9%)	641 (21.1%)
use: Math is useful in everyday problems	887 (29.6%)	2114 (70.4%)
logic: Math helps logical thinking	1068 (35.6%)	1930 (64.4%)
boybet: Boys are better at math than girls	2610 (86.9%)	395 (13.1%)
job: Math is needed for a good job	722 (24.0%)	2286 (76.0%)
often: I will use math often as an adult	791 (26.3%)	2219 (73.7%)

Table 1. Frequencies and percentages (in parentheses) of 12 dichotomous, Grade 7 math attitude items.

Four factors were extracted with an EFA for categorical outcomes using a weighted least squares with mean and variance adjustment (WLSMV) method of estimation. Though eigenvalues are greater than one for the first three components (3.778, 2.538, 1.504). 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 fourth factor has an eigenvalue of .906 and the screeplot indicates a break between component 4 and 5. It seems to start to level off between the 4th and 5th factor and in addition, the four factors provide a more interpretable solution.

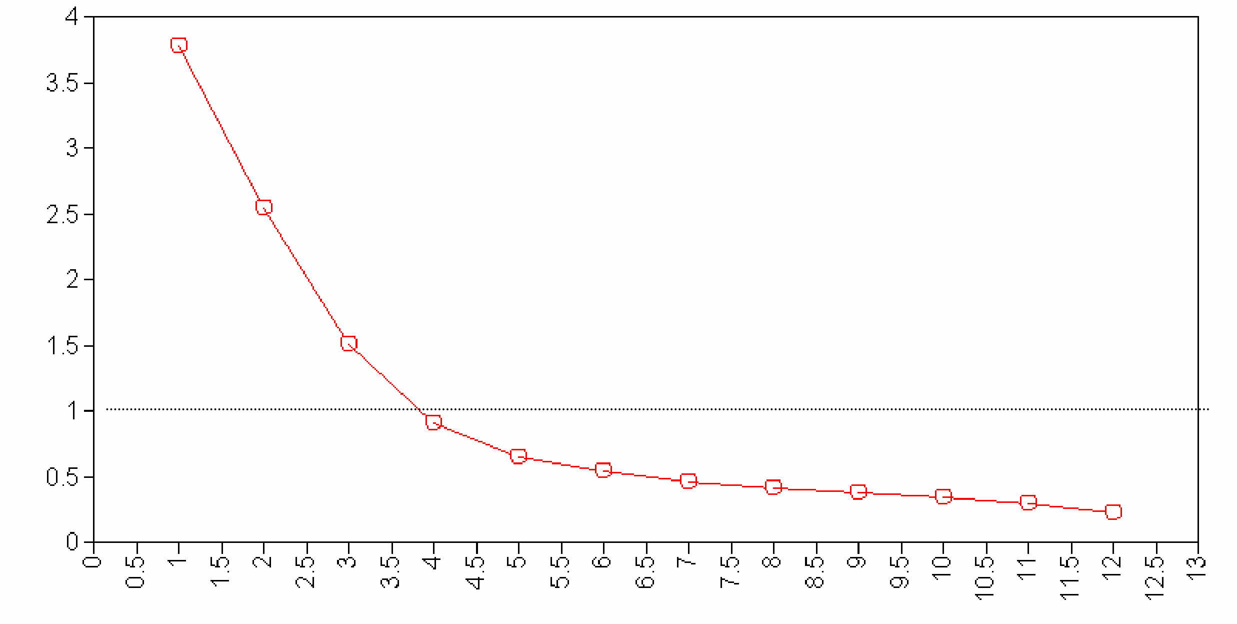

Figure 1. Eigenvalue plot for exploratory factor analysis.

A promax (oblique) rotation provides a conceptually clearer picture of the factors. Factor 1 includes items pertaining to self efficacy toward math (enj, good, und). These three items load highly on Factor 1 (>.6) and not as highly on the other 3 factors. Factor 2 includes items that deal with the value or use of mathematics (use, logic, job, often). These items load highly on Factor 4 (>.6) and less high on the other 3 factors. Factor 3 includes items that pertain to anxieties toward math (nerv, wor, scar). These items load high but negatively on Factor 2 (<-.5) and not as high on the other 3 factors. Factor 4 includes items that deal with beliefs about gender (useboy, boybet). These items load high but negatively on Factor 3 (<.8) and not as high on the other 3 factors.

Results from this factor analysis suggest four dimensions of math attitudes based on responses to these 12 items. These dimensions have to do with self efficacy, anxiety, gender beliefs and use. However, these results don't necessarily tell us whether there are groups of respondents who are similar in their responses to these twelve items. This is the advantage of conducting a latent class analysis. Factor analysis and latent class analysis are viewed as "complimentary" approaches[1]. The advantage to conducting a LCA versus a factor analysis is that "LCA helps find clusters of individuals who are similar, whereas this is difficult in factor analysis." LCA models the extent to which there are groups of individuals who are similar in terms of some unobserved heterogeneity. The comparability of these approaches will be discussed after discussing the interpretation from the LCA.

[1] Muthen, B. (2001). Latent variable mixture modeling. In G.A. Marcoulides & R.E. Schumacker (Eds.), New Development and Techniques in Structural Equation Modeling (pp. 1-33). Mahwah, N.J: Lawrence Erlbaum Associates.

Though the EFA suggested 4 factors, and a 5 class LCA solution probably fits best, I started with a 1 class solution and kept adding classes. Table 2 compares fit indices for the different number of classes.

Number of Classes	*Loglikelihood*	*Number of Parameters*	*BIC*	*AIC*	*Entrophy*	*LRT p-value for k-1*
1	-19997.189	12	40090.719	40018.377	N/A	N/A
2	-18752.096	25	37704.904	37554.192	0.711	0.0000
3	-18311.976	38	36929.034	36699.952	0.702	0.0000
4	-18063.093	51	36535.637	36228.186	0.758	0.0002
5	-17895.422	64	36304.665	35918.844	0.735	0.0003
6	-17830.108	77	36278.407	35814.216	0.731	0.0547
7	-17778.315	90	36279.190	35736.629	0.720	0.0226

Table 2. Summary of latent class analyses with 1-7 classes

The loglikelihood value is largest at 7 classes. The BIC and AIC are also lowest in the 7 class model which would suggest that the 7 class model fits the data well. However, in checking the loglikelihood values for the 10 best solutions out of 100 for the 6 and 7 class models, these values are very different from one another (i.e. -18045.822 and -17983.558 as extreme loglikelihood values for the 7 class model and -18115.867 and -17947.179 as extreme loglikelihood values for the 6 class model). This causes concern that there is an attempt to extract too much from the data and that a lower number of classes fits the data better. Thus a 5 class solution, in which the loglikelihood value at the local maxima for the 10 best solution is stable (-17895.422) and doesn't bounce around as the 6 and 7 class solutions seems plausible. The BIC and AIC values are lower in the 5 class model compared to the 1-4 class models. This suggests that a 5 class solution is an improvement from these models with a lower number of classes.

The entropy value is a measure of the amount of correct classification. There are similarities to an R^2 in that it is used to assess the usefulness of the model. This value should be high. The values for 2-7 classes are all greater than .7 but is the highest for 4 classes (.758). All of the values are relatively high so this value is not particularly helpful in determining the number of classes. It is helpful to know that all of the classes have high values but beyond that, this value is not very helpful for this particular exercise.

The LRT p-value for k-1 classes is the Vuong-Lo-Mendell-Rubin likelihood ratio test for k-1 classes. If this value is significant, there is evidence to suggest that the current model fits better than the previous model (k-1) with 1 less class. For example, in Table 2, the LRT p-value for the 4 class model which tests for 3 versus 4 classes is 0.0002. This is a significant value which suggests that the 4 class model is better than the 3 class model. We reject the 3 class model in favor of the 4 class model. Following this logic, the 6 class model is not significant (0.0547) which suggests that the 6 class model does not fit better than the 5 class model. We fail to reject the 5 class model for the 6 class model. Another guide is to look at the EFA which suggests 4 factors. In some situations, this would suggest a 5 class solution. Based on these reasons, I think a 5 class solution fits the data best.

Mplus provides different ways of determining the proportion of individuals in each class and the relative size of the class (how many respondents are in each class).

1. based on estimated posterior probabilities: Each respondent is assigned a probability for being classified in a particular class given their responses. For a given class, this method sums the probability of being in that class across all respondents.
2. based on most likely latent class membership: Each respondent is assigned a probability for being in a particular class, given their responses. These probabilities differ depending on the pattern of responses but sum to 1.0 for each respondent. This method finds the class that the respondent is most likely to be classified in (where the probability is the highest) and puts the respond in the class and then counts the number of respondents in each class and divides by the total number of respondents.

Table 3 provides the proportions of the sample classified into each class based on the different methods.

	est. posterior probabilities	*most likely latent class membership*
1	0.08661	0.08412
2	0.20907	0.20998
3	0.18437	0.18748
4	0.44064	0.44408
5	0.07930	0.07434

Table 3. Proportion of respondents classified in each of the 5 classes for each method.

Most of the respondents are classified in Class 4 (approximately 44%). There are few respondents in Class 5. The percents aren't exactly similar using the different methods of classification but the relative magnitude of the class sizes remains similar.

In addition to how many respondents are in each class, one might wonder how *well* the respondents are classified into each of these classes. Since each respondent has a particular probability for being in each class, an average of these probabilities for each respondent in each class is calculated by Mplus (Table 4).

	1	*2*	*3*	*4*	*5*
1	**0.839**	0.091	0.000	0.000	0.069
2	0.046	**0.803**	0.047	0.122	0.033
3	0.033	0.058	**0.774**	0.122	0.013
4	0.000	0.039	0.048	**0.892**	0.020
5	0.001	0.061	0.030	0.088	**0.819**

Table 4. Average latent class probabilities for the most likely latent class membership (row) by latent class (column).

If respondents were perfectly classified in a latent class, we would see a 1.000 in the diagonal and a 0.000 in the off-diagonals. A 1.00 would indicate perfect classification in a particular class and absolutely no chance of being classified in another class. Higher number indicates that based on their probability for being in that class, respondents were in fact classified in that particular class. Table 4 indicates that the diagonals are the large (0.839, 0.803, 0.774, 0.892, 0.819) and the off diagonals are small. This is evidence that respondents were well classified into these 5 classes.

Table 5 provides the probability of each class endorsing a particular item for the 4 and 5 class models. It also includes the prevalence as defined as the proportion of respondents for the latent class patterns based on the estimated posterior probabilities.

	4-Class Solution[1]				5-Class Solution[2]				
	1	2	3	4	1	2	3	4	5
Prevalence	*0.190*	*0.191*	*0.537*	*0.082*	*0.087*	*0.209*	*0.184*	*0.441*	*0.079*
enjoy	0.457	0.259	0.893	0.900	0.132	0.309	0.763	0.903	0.902
good	0.497	0.200	0.915	0.907	0.079	0.243	0.852	0.924	0.919
und	0.526	0.452	0.954	0.876	0.175	0.496	0.835	0.963	0.884
useboy	0.122	0.128	0.039	0.663	0.162	0.118	0.091	0.036	0.715
nerv	0.237	0.535	0.066	0.477	0.354	0.516	0.126	0.065	0.573
wor	0.521	0.831	0.631	0.772	0.457	0.845	0.569	0.636	0.764
scar	0.213	0.387	0.077	0.616	0.316	0.378	0.124	0.075	0.623
use	0.187	0.737	0.847	0.886	0.198	0.744	0.323	0.907	0.899
logic	0.195	0.610	0.784	0.844	0.148	0.631	0.326	0.843	0.847
boybet	0.118	0.127	0.052	0.645	0.121	0.126	0.114	0.043	0.696
job	0.247	0.880	0.879	0.887	0.222	0.880	0.423	0.924	0.891
often	0.183	0.775	0.893	0.908	0.121	0.801	0.379	0.942	0.920

Table 5. Latent class analysis of LSAY math attitudes for Grade 7 (n=3067).

[1] Likelihood ratio chi-square fit=-18063.093, with 51 degrees of freedom
[2] Likelihood ratio chi-square fit=-17895.422, with 64 degrees of freedom

Information for the 5 class solution is also graphically represented in Figure 2. To assist with the interpretation of the factors, the items were rearranged so that items that loaded highly on the same factors, based on the EFA, are located next to one another.

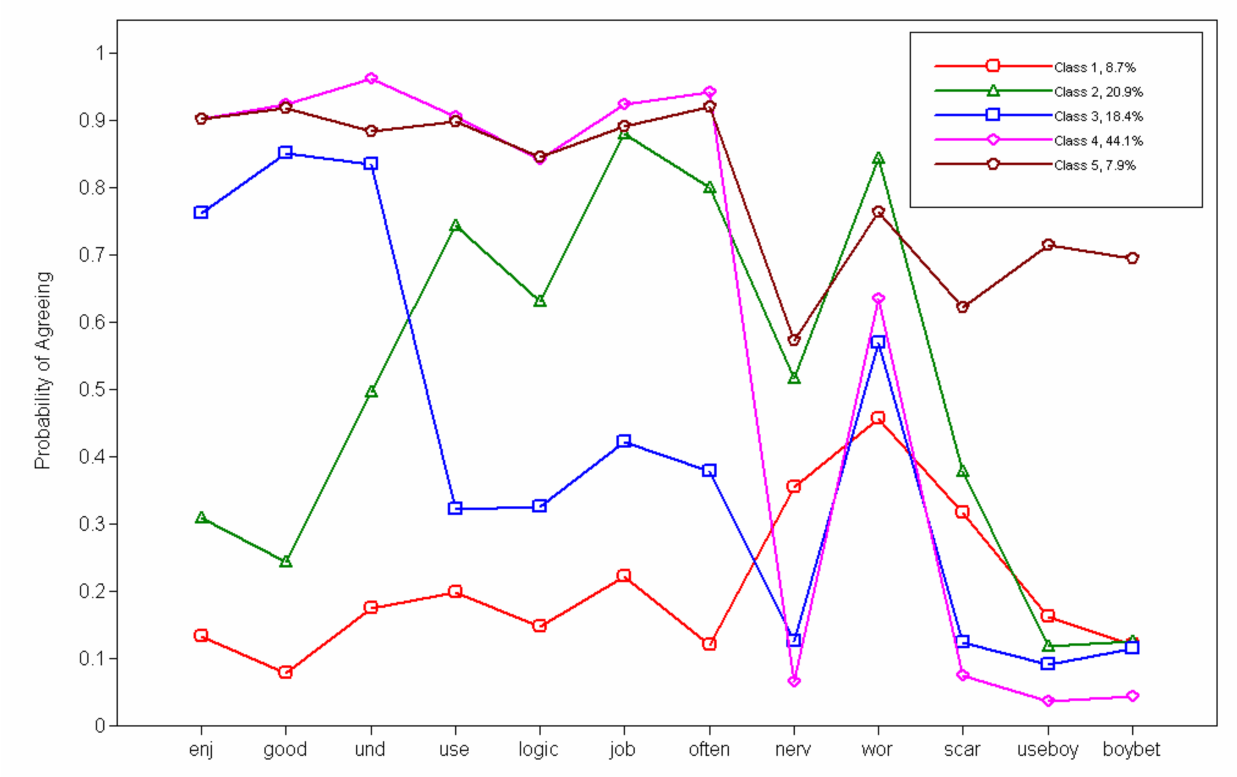

Figure 2. Probability of agreeing with each item

The next step is the interpretation of the classes. To help with these interpretations, I merged the information about what classes the respondents were classified in with achievement data (Grade 7 and Grade 12 performance in algebra, geometry, quantitative literacy) and dropout intentions (dropot: have you ever thought about dropping out in Grade 7; drop7-12: dropout intentions in Spring of Grades 7-12).

Respondents in Class 1 do not agree with most of these items. I would consider this class to be anti-math and have a general disinterest or frustration toward math. Approximately 9% are classified into this class in which there are relatively low proportions of these respondents who endorse the 12 items. These respondents tend to disagree that the enjoy math, are good at math or understand mathematics. These respondents might likely be a group that you might consider to have a low motivation or interest in mathematics and might be in danger of dropping out. In fact, comparing their thoughts of dropping out from Grades 7-12 indicates that this group of individuals does tend to have higher thoughts of dropping out compared to the other classes (approximately 12%). Whether they actually dropout might be something altogether different but the fact that they are willing to admit that that they have thoughts of dropping out and don’t seem to be performing as highly in math achievement measures suggests that this group is probably frustrated with math and aren’t doing much to change these negative feelings toward math. I wouldn’t be surprised if most of these students are in lower level mathematics classes and might not be around other students who have less interest in math. Even the items that might be considered negatively scored (wor7, scar7) could be interpreted as a disinterest. So, even though the respondents did not agree that they were worried about math test grades or were scared when opening a math book, they might just be disinterested or not care about math or school in general and thus might considered a group of students who are disinterested in math and don't see much of a future needing math skills or knowledge.

Respondents in Class 2 are the high anxiety, high achieving group. This group sees value or use in math in the future. They don't buy into the gender stereotypes that boys are better than math than girls or that boys will use math in the future more than girls. This group also has a high anxiety toward math by tending to agree with items that deal with being nervous toward math, worrying about math test grades, and being scared when opening a math book. It might be that this group, despite their high performance in Grade 7 and Grade 12 on various math achievement tests, also tend to be anxious and worry about their performance. These students do not have much thoughts of dropping out (the lowest percentage from Grades 7-12 compared to the other classes of students) but still have anxieties about math. These students are probably the good kids who have some initial low attitudes toward math despite being high performers.

Respondents in Class 3 are contentious, high achievers that don't value math or see the utility of math in the future. This group tended to endorse the self-efficacy items indicating that they tended to enjoy math, thought they were good at math and usually understand math. They also tended to have high Grade 7 math achievement scores. However, respondents also tended to see the value and use of math. They also disagreed that they had anxieties toward mathematics but did tend to worry about their math test grades which suggests that they are a contentious group. They weren't scared of opening math books or nervous about math but they do seem to be anxious about their performance on math tests. These respondents also disagreed that boys were better at math or math was more useful for boys which suggests that they don't buy into the stereotype that boys are better at math than girls. This progressive attitude seems to translate to caring about their performance in math but also feeling confident in their own personal abilities without paying much attention to stereotypes. This group of respondents has the lowest thoughts of dropouts throughout middle and high school compared to the other classes of students. These respondents also had high Grade 12 achievement test scores compared to the other groups which is why I think respondents in this class tend to be contentious and high achieving in math.

Respondents in Class 4 is similar to Class 3 in that they are contentious. However they aren't as high achieving but see value and use to math. These respondents tend to endorse the items pertaining to self efficacy (enj7, good7, und7). These respondents are also contentious in their math test performance because they tend to endorse the item pertaining to being worried about their math grades. These respondents are also secure in their abilities and do not seem to buy into the gender stereotypes of boys being better in math than girls. The primary difference between respondents in this Class 3 and Class 4 is that respondents in Class 3 do not see much value or use for math in the future. They don't feel as though they need math for a better job or that they will use math often as an adult. They are still high achievers and probably a "good" group of students and don't think about dropping out compared to the other groups. However, they see the value or use of math which makes them different from respondents in Class 3.

Respondents in Class 5, are the blissfully enthusiastic group. These respondents had high probabilities of agreeing with all of the items. They tended to agree that they enjoyed math and were good at it. These respondents also reported that they were less likely to be nervous, scared or worried about math. But in looking at the average Grade 7 measure of math achievement in algebra, geometry, quantitative literacy and general mathematics, this group of respondents actually tends to be pretty low compared to the other four classes. They seem to have an awfully high self esteem and attitude toward math even when they don't have much reason to feel good about their performance in math. This is well and fine but what I consider blind enthusiasm and self delusion. Another theory is that these respondents maybe didn't take the survey items very seriously and that they were probably not really paying much attention to the items. It seems unlikely that this group of respondents who have some of the lowest average scores as compared to the other classes could think that they are good at math. I think this group of respondents is just overly enthusiastic or not very good judges of their ability. It might also be that this group of respondents might not care very much in the way that they answered the items which puts them at high probabilities of endorsing everything even though it doesn't seem to be validated by other evidence. This class has the highest percent of students having thoughts of dropping out. On average, these students also tend to be the lowest performing on algebra, geometry, quantitative literacy and math achievement items in Grades 7 and Grade 12. Whatever the case, this group of respondents seems to agree and endorse just about every item but really not have much cause for their enthusiasm.

The 5 latent classes were plotted for Factor 1 (self efficacy toward math: enjoy, good understand) and Factor 2 (use and value of math: use, job, often, logic). Figure 3 shows the factor scores for Factor 1 and Factor 2 by latent class.

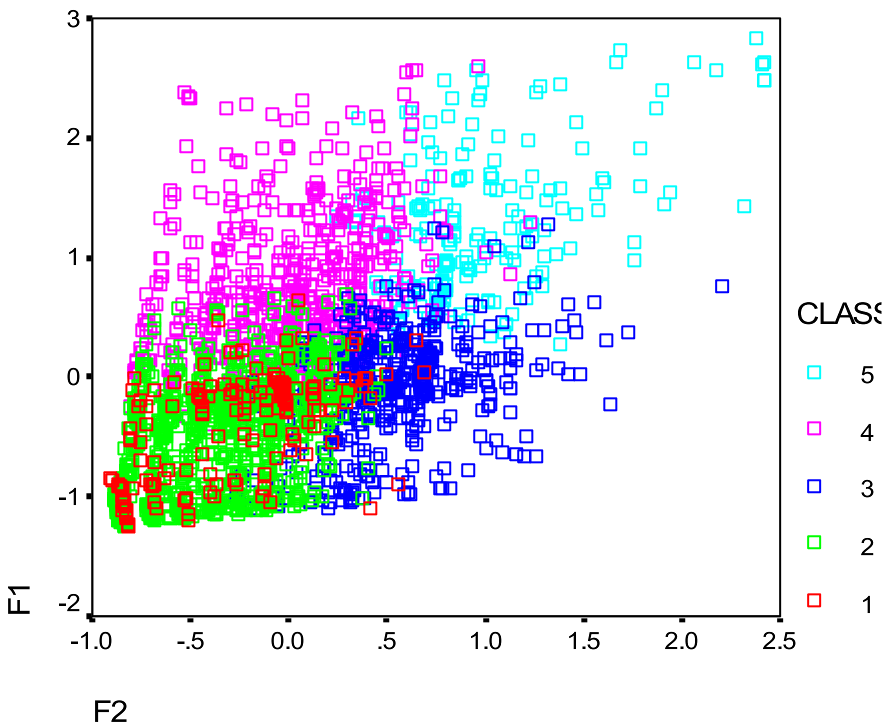

Figure 3. Factor 1 and Factor 2 scores by latent class.

Figure 3 indicates that isn't a clear cut point between the factor scores. Latent class helps to find these cut points. From the Figure, there seems to be differences between the classes in terms of their factor scores on these two dimensions of math attitudes. The EFA and LCA are complimentary in that information from each analysis was used to help inform the other. The EFA suggested 4 factors and provided information about how the items were grouped in these four factors. This was helped to order the items in the plots for the LCA. I grouped the items that were on the same factor together which helped to make the plots interpretable.

I attempted a few covariates that deal with home resources, mothers level of education, fathers level of education, mothers social economic index, fathers social economic index, gender and Grade 7 ability ranking. Class 5 is my comparison class. The start values for the 5 class model were used to order the classes. Below is a table summarizing the estimates for the categorical latent variables for female, Grade 7 ability ranking and home resources.

	Estimates	*S.E.*	*Est./S.E.*
C#1 on			
Female	1.576	0.351	4.485*
Ability ranking	-0.124	0.097	-1.278
Home resources	0.089	0.067	1.323
C#2			
Female	2.089	0.320	6.524*
Ability ranking	-0.073	0.100	-0.732
Home resources	0.163	0.067	2.423*
C#3			
Female	1.897	0.326	5.811*
Ability ranking	0.030	0.102	0.292
Home resources	0.148	0.067	2.214*
C#4			
Female	1.895	0.314	6.039*
Ability ranking	0.189	0.090	2.109
Home resources	0.258	0.058	4.448*

Table 6. Estimates for covariates for 5 class, with Class 5 being the comparative class.

These estimates can be transformed into odds ratio by taking the exponential of the estimate. For example, females are 4.8% ($e^{1.576}$) times more likely to be in Class 1 versus Class 5 holding all other covariates constant. Females are generally more likely to be in all classes except Class 5. Class 5 is what I labeled the blissfully enthusiastic group. It might suggest that males are more likely to be over confident or not want to show that they don't understand math or don't like math. Males may also be frustrated or disillusioned with math and might not have taken the items very serious and thus filled everything positively even though they had no evidence to show that they were actually good at math.

Gender is a significant covariate for all classes. The math ability group that students were placed for Grade 7 seems to be a significant covariate for only Class 4. This seems to indicate that math placement in Grade 7 only important consideration in this fourth class. Home resources only seems to be a significant covariate for Class 2, 3 and 4. It is not a significant covariate for Class 1. Another way to interpret these results is to look at the plots. Below is an example of plots for comparing placement in different ability levels for females and males with mothers with less than a high school degree.

Females Males

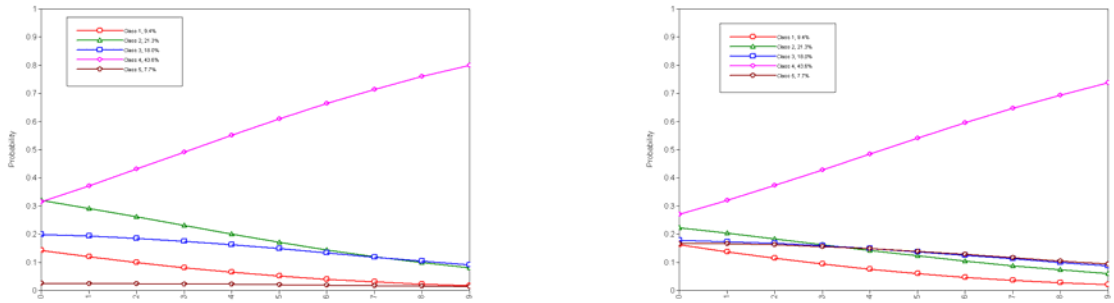

Figure 4. Distribution of home resources by class membership for females and males who have mothers with less than a high school education.

The plots indicate that males and females are most likely to be in Class 4. The plots provide a distribution of the classes across the levels of home resources conditional on gender and mothers education. These types of plots are helpful in looking for even more patterns among the classes and help with the interpretation of class membership.