Assignment 6

The purpose of this assignment is to explore two methods of modeling change: growth mixture modeling (GMM) and latent class growth analysis (LCGA). The difference between GMM and LCGA is that GMM allows for within-class variation in the latent trajectory classes and LCGA does not allow for within-class variation. In other words, LCGA does not allow for individual differences within the latent classes.[1] Since GMM considers separate growth models for each class, this allows for differences in the effects of the covariates, residuals, and growth functions. This might result in differences in terms of growth rates for the different classes or differences in terms of the significant effects and magnitude of the effects of the covariates in the different classes.[2]

Mplus Version 3.01 was used to carry out a GMM and LCGA on the LSAY data set math achievement scores for Grades 7-12. Two assigned covariates (gender and educational expectations) were also used. The initial sample size included 3116 students. However, 14 cases were missing data on all of the achievement outcomes and 40 cases were missing data on the covariates. Models without covariates included 3102 cases and models with covariate(s) included 3062 cases.

Table 1 presents descriptives for the math achievement variables (mth7, mth8, mth9, mth10, mth11, mth12) for Grades 7-12 used in these analyses. The sample size decreases substantially from Grades 7-12 so all cases are used in the analyses. There seems to be a substantial loss in Grades 11 and 12, which is why listwise deletion was not used in these analyses. Instead, cases are assumed to be missing at random. There is a positive increase in math scores as well as an increase in the variation of scores from Grades 7 to Grade 12.

	N	*Mean*	*Std. Dev.*	*Variance*	*Minimum*	*Maximum*
Grade 7	3065	50.39	10.20	104.02	27.28	85.02
Grade 8	2581	53.83	11.02	121.51	24.92	88.54
Grade 9	2241	58.81	12.60	158.87	26.57	94.19
Grade 10	2040	63.57	13.65	186.30	29.60	95.17
Grade 11	1593	67.64	13.77	189.72	31.46	97.29
Grade 12	1168	68.61	14.79	218.79	27.31	98.36

Table 1. Descriptive information of the math IRT scores for Grades 7-12.

Table 2 presents correlations for these achievement variables. There are fairly high correlations between these achievement variables (>.75) which suggest that achievement over time is highly related.

[1] Muthen, L.K. and Muthen, B.O. (1998-2004). *Mplus User's Guide. Third Edition*. Los Angeles, CA: Muthen & Muthen.
[2] Muthen, B.O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences.* Newbury Park, CA: Sage Publications.

	Grade 7	*Grade 8*	*Grade 9*	*Grade 10*	*Grade 11*	*Grade 12*
Grade 7	- - -					
Grade 8	.830 (n=2550)	- - -				
Grade 9	.806 (n=2214)	.867 (n=2037)	- - -			
Grade 10	.787 (n=2019)	.831 (n=1837)	.902 (n=1821)	- - -		
Grade 11	.773 (n=1574)	.812 (n=1438)	.867 (n=1410)	.910 (n=1430)	- - -	
Grade 12	.766 (n=1156)	.810 (n=1065)	.855 (n=1033)	.892 (n=1040)	.923 (n=977)	- - -

Table 2. Correlation between math IRT achievement scores. All correlations are significant at the .01 level (2-tailed).

There are 1591 males and 1471 females included in this sample. Tables 3 provides frequencies for the two covariates (female, expect). The student expectation covariate is collected in Grade 7 and is coded as follows: 1=HS only, 2=Vocational training, 3=some college, 4=Bachelor's, 5=Master's, 6=Dr, PhD. Only 44 cases are missing information on this variable. A majority of the respondents expect to obtain at least a Bachelor's degree. Approximately 30% of the students in this sample expect less than a Bachelor's degree.

	Male	*Female*	*Total*
HS only	135	81	216
Vocational training	137	70	207
Some college	301	256	557
Bachelor's degree	470	447	917
Master's degree	307	333	640
Dr, Ph.D.	241	284	525
Total	1591	1471	3062

Table 3. Frequencies of student's educational expectations by gender.

Latent Class Growth Analysis (LCGA)

LCGA is considered a "special type of growth mixture model" in that "individuals within classes are treated as homogenous with respect to their development." To accomplish this, these LCGA analyses are carried out with zero growth factor variances and covariances.[3] There were 3102 cases used in these analyses (14 cases missing data on all variables). Table 5 presents fit indices for 1-7 classes for the LCGA.

Number of Classes	*Loglikelihood*	*Number of Parameters*	*BIC*	*AIC*	*Entrophy*	*LRT p-value for k-1*
1	-49673.928	8	99412.174	99363.856	N/A	N/A
2	-46041.720	11	92171.878	92105.440	0.842	0.0000
3	-44247.873	14	88608.304	88523.747	0.850	0.0000
4	-43473.400	17	87083.476	86980.799	0.840	0.0000
5	-43071.706	20	86304.207	86183.411	0.812	0.0236
6	-42853.512	23	85891.940	85753.024	0.785	0.0002
7	-42762.914	26	85734.864	85577.829	0.761	0.2108

Table 4. Fit indices for latent class growth analysis for 1-7 classes and no covariates.

The entropy values are fairly high for the different number of classes (>.76). The BIC and AIC values continue to decrease as the number of classes increase. There is no dip at which point the values start to increase so using these indices aren't very helpful in this example for determining the number of classes. The likelihood ratio test is not significant at 7 classes which would suggest that the 6 class model is sufficient. But it is not clear from these indices what number of classes is the best. Checking the proportion of students classified into the 6 class solution and 5 class solution, it seems like the additional class in the 6 class model is a combination of several classes. It seems to be students who started off moderate to low and grew at a moderate rate. Turning to substantive questions or hypotheses and trying to interpret the different number of classes and running a GMM would be the next steps. Because the LCGA does not allow for class specific variation, I would expect this LCGA modeling approach to include a higher number of classes to help explain some of the variation. The mean intercept and slope values for each class were used as starting values in the GMM models to follow.

[3] Muthen, B.O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of Quantitative Methodology for the Social Sciences.* Newbury Park, CA: Sage Publications.

Growth Mixture Model (GMM)

In the GMM framework, differences in variances between the latent classes are permitted. The first strategy was to conduct a GMM without using the 2 assigned covariates. I started off with the default setting in Mplus described in the Mplus User's Guide Example 8.1 which allows the growth factor intercepts to be different across classes but holds the residual variances and residual covariances of the growth factors equal across classes. A maximum likelihood estimator with robust standard errors using a numerical integration algorithm was used for these models (algorithm = integration).

Number of Classes	*Loglikelihood*	*Number of Parameters*	*BIC*	*AIC*	*Entrophy*	*LRT p-value for k-1*
1	-42711.235	11	85510.908	85444.470	N/A	N/A
2	-42639.400	14	85391.357	85306.800	0.574	0.0000
3	-42576.788	17	85290.252	85187.575	0.675	0.0000
4	-42520.275	20	85201.347	85080.551	0.626	0.0000
5	-42502.467	23	85189.850	85050.935	0.594	0.3664
6	-42502.493	26	85214.021	85056.986	0.620	0.6853
7	-42478.375	29	85189.905	85014.751	0.635	0.4911

Table 5. Summary of fit indices for a growth mixture model without class specific variances and without covariates.

The likelihood ratio test suggests that a 4 class solution fits the data best. The BIC and AIC dips at 5 classes which suggests that the 5 class solution fits the model best. But the Vuong-Lo-Mendell-Rubin likelihood ratio test suggests a 4 class solution fits best. In looking at the 5 class solution, one of the classes has a low number of students classified in this particular class (4%). This class of students seems to start off moderately high but has a high rate of growth. However, the entropy value for the 5 class solution is lower (0.594) which suggests that students are not classified well. The students in the small class have a 12% probability of being classified with students who have similar initial starting values to another class. The probability of being classified in another category seems fairly high for all of the classes in this model, which is why the entropy value is lower than the other models. The entropy value for the 4 class solution is higher than the 5 class solution so there seems to be some justification for the 4 class solution, despite lower BIC and AIC values in the 5 class solution.

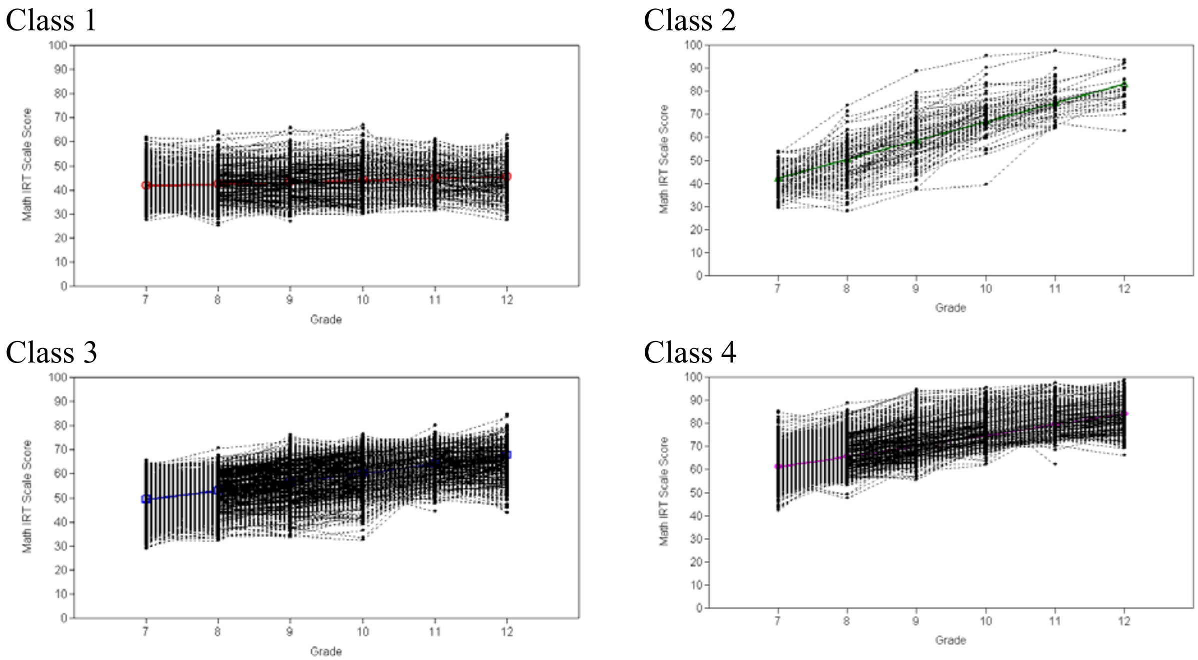

Figure 1. Estimated means and observed individual trajectories for 4 class growth mixture model with no variation between the classes and no covariates.

In trying to refine the 4 class model, I checked for class specific variance by looking at the plots of the estimates for the variances of the average initial starting point and growth factor means for each class (Figure 1). Class 2 seems to have a different intercept and slope variation than the other classes which suggests that Class 2 might have a class specific variance. I continued with the 4 class solution and attempted to allow different things to vary (Table 7).

	Loglikelihood	*-2*Difference*
Class invariance	-42520.275 (20)	
1. Class 2 intercept variance different from Class 1, 3, & 4	-42493.570 (21)	53.41 (1)*
2. Class 2 slope variance different from Class 1, 3, & 4	-42507.067 (21)	26.42 (1)*
3. Class 2 intercept and slope variances different from Class 1, 3, & 4	-42506.321 (22)	27.91 (2)*
4. Class 2 slope with intercept different from Class 1, 3, & 4	-42518.249 (21)	4.052 (1)*
5. Class 2 outcome residuals different from Class 1, 3, & 4	-42307.315 (26)	425.92 (5)*
1 and 5. Class 2 intercept variance and outcome residual variances different from Class 1, 3, & 4	-42291.639 (27)	457.27 (7)*

Table 6. Loglikelihood values and difference between loglikelihood values. The difference is between the class invariant model and the Class 2 specific variance. The difference between the two models is then multiplied by -2. This difference is chi-square distributed. These are 4 class growth mixture models. * indicates a significant improvement.

In looking at the chi-square difference test from the invariant model (variance equal across classes), there seems to be improvements when allowing Class 2 to have different intercept variance, slope variance, intercept and slope variance, slope with intercept variance, outcome residual variances when compared to Class 1, 3 and 4. Though, it seems logical that all of these things could be class specific, I opted to just select the two changes that seemed to produce the largest loglikelihood differences: intercept variances and outcome residuals which are different for Class 2 than Class 1, 3 and 4.

A further step to improve the model is to add covariates. Two covariates: female and expect were used in this assignment. Sometimes the addition of covariates changes the class formation relative to the class specific variance without covariates. To check this, I ran a 2, 3, 4 and 5 class GMM solution with the two covariates to see if the fit indices would continue to support a 4 class solution (Table 8).

Number of Classes	*Loglikelihood*	*Number of Parameters*	*BIC*	*AIC*	*Entrophy*	*LRT p-value for k-1*
2	-41835.553	20	83831.643	83711.107	0.608	0.0000
3	-41766.355	25	83733.380	83582.710	0.603	0.0000
4	-41710.308	30	83661.421	83480.617	0.625	0.0000
5	-41682.801	35	83646.540	83435.601	0.625	0.1182

Table 7. Summary of fit indices for growth mixture model with class invariance and gender and educational expectations as covariates.

The inclusion of the two covariates continues to provide support for a 4 class solution. The entropy value is similar to the 4 class model without covariates. The likelihood ratio test is not significant for the 5 class model which suggests that the 4 class model is sufficient. The AIC and BIC values are decreasing but do not seem to dip.

The estimates suggest that there is no significant effect of gender. There seems to be a significant impact of educational expectations but gender does not seem to be very informative in terms of looking at where students start off and how fast they grow. So, I decided to exclude gender as a covariate and instead include only educational expectation.

The default in GMM is for class invariance in terms of the growth factors on the covariate. This default does not allow for differential effects of educational expectations for each class. Again, one way to check for this is to look at the plots for the four class solution.

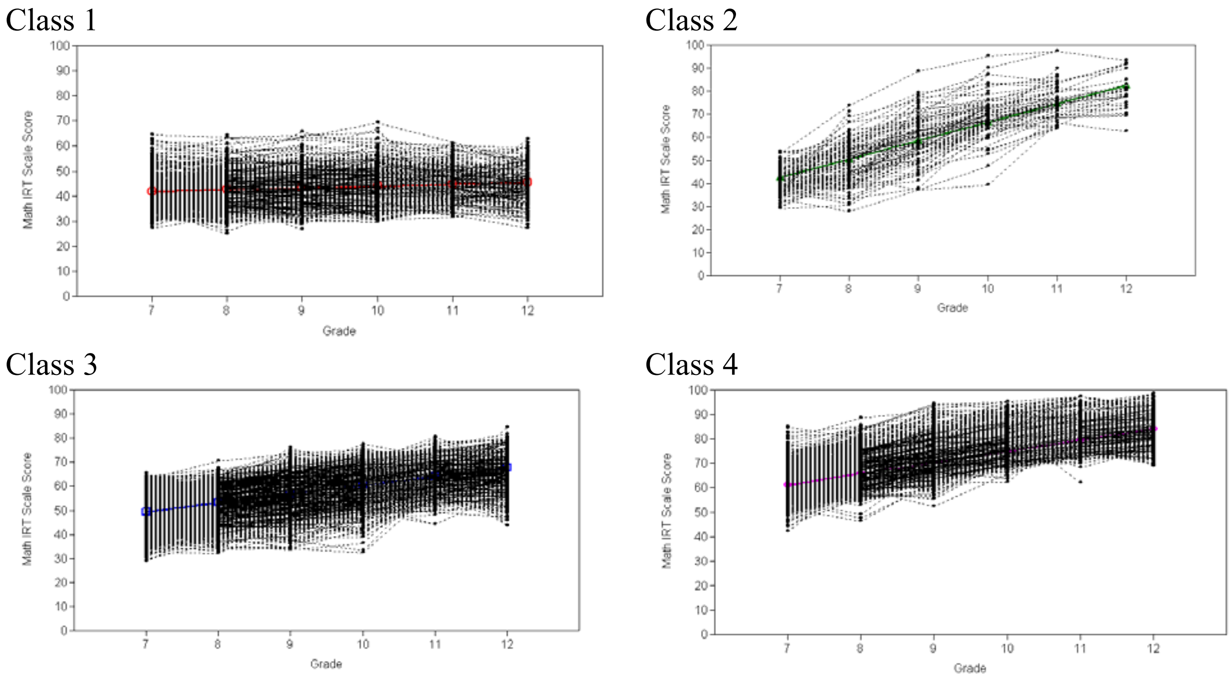

Figure 2. Estimated means and observed individual trajectories for 4 class growth mixture model with no variation between the classes and 1 covariate.

The plots suggest that Class 2 seems to have variances that might be different than the other 3 classes. Table 9 compares the fit indices for the 4 class invariant model with 1 covariate. The first run uses the default settings of equal variances across all classes, and equal effect of the covariate across classes. This is what I considered the default of baseline model.

The next step was to allow the class variances to differ. From rationale earlier presented, the intercept variances and the outcome residuals for Class 2 is allowed to be different from the intercept variances and outcome residuals for Class 1, 3 and 4. The effect of the covariate is similar in this model. There is an improvement in this model over the default model of class invariance. However, the entropy value is lower in this model (0.571).

The next step was to hold the class variances similar but allow the effect of the covariate to differ. This allows the effect of educational expectations to differ for Class 2 but to be held constant for Class 1, 3 and 4. This seems to be the best model and will be interpreted at the end of this assignment. There is no class specific variance in terms of the intercept or outcome residuals for Class 2 in this model.

The final step was to allow the class variances to differ and the effect of the covariate to differ. The addition of the covariate *and* allowing for class invariance in Class 2 *and* allowing for class specific effects of the covariate does not seem to improve the model. The inclusion of the covariates seems to take away with the need for class invariance.

The 4 class solution without covariates also seems to fit well, but substantively, I could see an argument for including educational expectations as ways of talking about the differences between classes.

Equal effect of covariate	*Equal class specific variance*	*Loglikelihood*	*Number of Parameters*	*BIC*	*AIC*	*Entrophy*	*LRT p-value for k-1*
Yes	Yes	-41738.080	25	83676.830	83526.160	0.629	0.0000
Yes	No	-41519.311	32	83295.481	83102.622	0.571	0.0008
No	Yes	-41737.674	27	83692.072	83529.348	0.629	0.0005
No	No	-41514.989	34	83302.889	83097.977	0.565	0.0004

Table 8. Summary of fit indices for 4-class growth mixture model with 1 covariate. The first model is the default that has equal effect of the covariate and equal class variances. The second model allows for class invariance in the intercept and outcome residuals in Class 2. The third model does not allow for class specific variance but allows for differences in the effect of the covariate in Class 2.

The model that I selected to interpret for this assignment has class specific effects of the covariate. In other words, the effect of educational expectations is presumed to be different for Class 2 than for Class 1, 3 and 4. I hypothesize that because Class 2 seems to be the high growth group, despite low initial performance, that maybe educational aspirations for this group was particularly different or functioned in a way that might be different from the groups who were low achievers, or were already high achievers in Grade 7.

The model that I selected to interpret for this assignment does not have class specific intercept variance or class specific residuals of mth7-mth12. Without covariates, the model that allowed for class specific variance seemed to be the way to go. However, with the addition of the covariate, it seems as though some of this variation disappears and is not necessary to specify. The model that I have selected to interpret allows for the variances in the growth factors to be similar across the classes. However, the impact of the covariate, educational expectations, is presumed to be different for Class 2. Class 2 is considered special in that though they were low performers in Grade 7, as a group they should tremendous growth and high performance by the end of Grade 12.

Interpreting the 4-class GMM with 1 covariate (specific effect of covariates in Class 2)

Table 10 provides the proportions of the sample classified into each class based on the different methods.

	est. posterior probabilities	*most likely latent class membership*
1	0.21328	0.20085
2	0.04423	0.02907
3	0.45614	0.49379
4	0.28635	0.27629

Table 9. Proportion of respondents classified into each of the 4 classes based on their estimated posterior probabilities and most likely class membership.

Most of the respondents are classified in Class 3 (46%). The proportion of respondents classified using each method seems roughly similar which is a good sign that respondents are consistently classified. Another thing to consider is to look at the probability of the most likely class membership (row) by the latent class that the student was placed (column). The diagonals are high and the off-diagonals are low which indicates that students were well classified in the most likely class and didn't have as high probabilities of being placed in another class (Table 11).

	1	*2*	*3*	*4*
1	**0.824**	0.014	0.161	0.001
2	0.001	**0.751**	0.150	0.098
3	0.096	0.031	**0.754**	0.120
4	0.002	0.016	0.171	**0.812**

Table 10. Average latent class probabilities for the most likely latent class membership (row) by latent class (column).

Figure 3 provides the estimated means of the 4 classes and the proportion of students in each class. There is a group of students that starts of low but seems to grow at a high rate and catches up to the students who started off high. Next, these mean trajectories will be explored further by looking at each class.

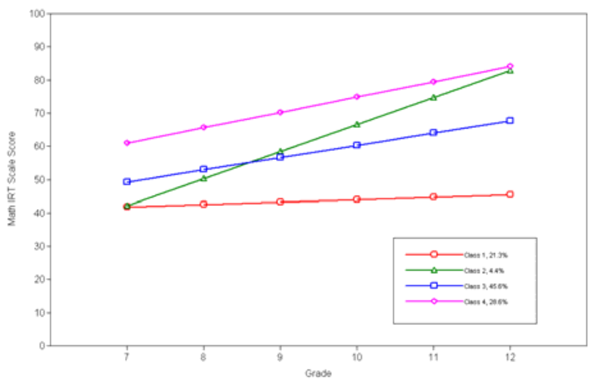

Figure 3. Estimated means for the 4 class GMM with 1 covariate (expect), effect of covariate different for Class 2.

There are approximately 21% of the students classified in Class 1. This class seems to start off low and not change very much in their math achievement scores from Grade 7 to Grade 12 (Figure 3). This group has an average start point of 38.189 and a 0.680 average rate of change. These estimates suggest consistent low performance and possibly disengagement or disinterest in school. I would suspect that this class might have a higher number of students who dropout of high school. For these low achievers, educational expectations is not significantly related to growth but is significantly related to initial starting point. It is somewhat promising to note that low educational aspirations is significantly related to Grade 7 performance but not necessarily related to change from Grade 7 through Grade 12. But this lack of relationship between growth and educational expectation might have something to do with there being so little growth to model.

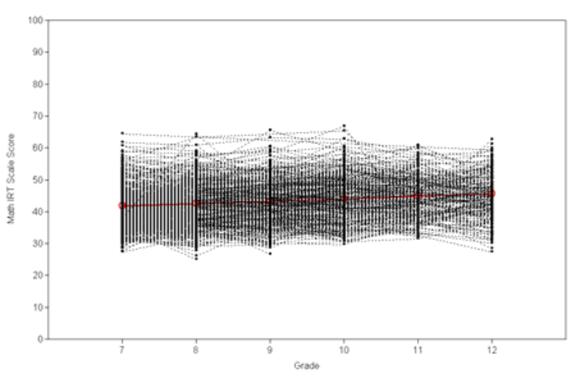

Figure 4. Estimated mean and observed individual trajectories for Class 1 (21% of respondents).

There are approximately 5% of the students classified in Class 2 (Figure 5). These students start off low but have a large rate of change. On average, students in this group have a Grade 7 math achievement score of 39.559. These students are just slightly higher than the low achievers in Class 1. However, these students are different from Class 1 in that they have a high average rate of change (8.357). These students might have started low but for some reason were able to show great gains in performance as they progressed through high school. This is interesting to note and might be worth figuring out what sorts of classes these students took or what sorts of teachers or curriculum that they were engaged in that made their growth much higher than the students who started off at the similar levels but did not grow as much. The impact of educational expectations varied for this particular group of students. For this particular group of students, their educational expectations were not related to their initial starting point or their large change in achievement. This is somewhat surprising because given the large growth displayed; I would suspect that educational expectations might have something to do with their growth. Obviously, there are other factors that might be more related to this growth that are not included in the model –such as home resources or maybe the educational aspirations of friends. Another option is the educational expectation variable was not well developed and did not measure the concept accurately or well.

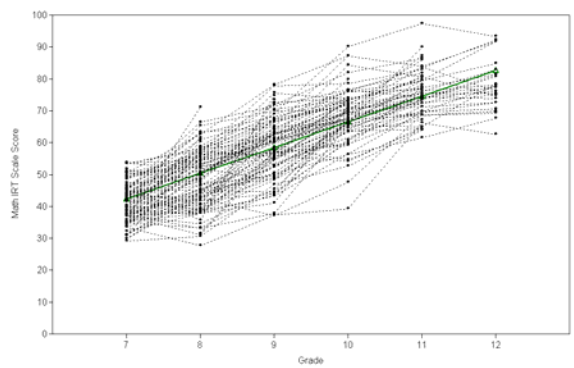

Figure 5. Estimated mean and observed individual trajectories for Class 2 (5% of respondents).

There are approximately 46% of the students classified in Class 3. On average, these students start off higher (45.188) in Grade 7 than students in Class 2. However, these students do not grow as much as students in Class 2 (3.628). These students might be the average students who aren't necessarily considered high risk for deviant behavior or persistent low performance but aren't considered high achievers either. As noted earlier, the impact of educational expectations was held constant for this particular class of students and it seems as though there is a significant impact of educational expectations on the intercept or average initial starting point but not a significant impact on growth or change over time.

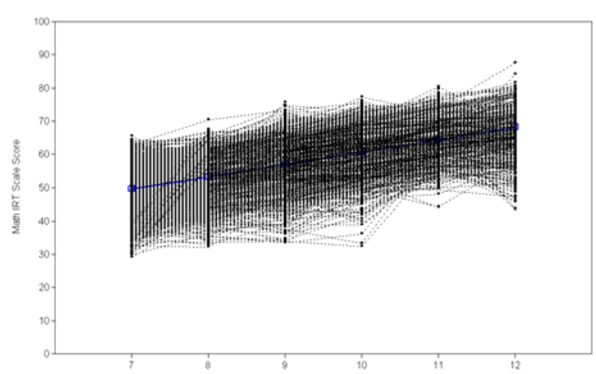

Figure 6. Estimated mean and observed individual trajectories for Class 3 (46% of respondents).

There are 28% of the students classified in Class 4. I would consider this group of students to be consistent high achievers. On average, this group of students start off higher than all other groups of students (56.010) and tend to grow at a moderate rate (4.553). They don't grow as much as Class 2 students but they start off higher than all other students. On average, students in Class 2 "catch up" with these high achievers in Class 4. Because these are high achievers, I would not expect there to be much high school dropouts in this class.

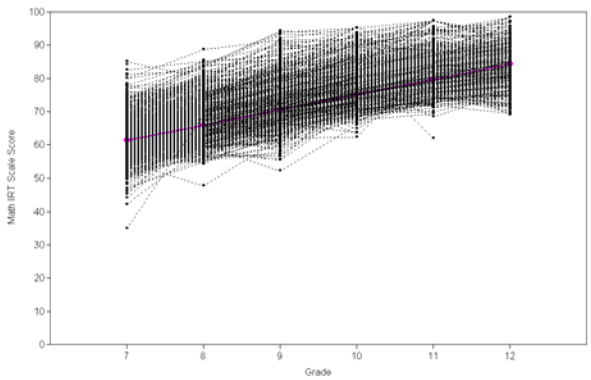

Figure 7. Estimated mean and observed individual trajectories for Class 4 (28% of respondents).

To determine whether class membership relates to high school dropout (lloctn=4), I ran the model using high school dropout as a distal outcome and looked at the frequencies of students within each class who were classified as a dropout.

Class	*LLOCTN=4 Dropped out of school*	*Total*
1: started low, low growth	189 (31%)	615
2: started low, high growth	7 (8%)	89
3: started medium, moderate growth	218 (14%)	1512
4: started high, moderate growth	30 (4%)	846
Total	444	3062

Table 11. Number of students who dropped out by 12th grade in each of the 4 classes.

This variable seems to be predictive of membership in Class 1. As suspected, students who were classified into Class 1 which is the low achieving and low growth class, had largest number (and percent) of students dropping out (Table 12). There were 189 students (31% of those student classified in Class 1) who dropped out by the 12th grade. There were low numbers of students who were classified in Class 2 and 4 who dropped out. This is not surprising given that Class 2 started off low but had high growth and Class 4 started off high and had moderate growth. The high achievers (even if they started off low) do not tend to drop out. Students who started off moderately high and had moderate growth dropped out less than students in the low achieving or failing class but not quite as low as students who were classified in Class 2 or 4.