Title
Moderation with a latent class variable: An applied example

Author
Arch, Dina

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

Moderation with a latent class variable: An applied example

A Thesis submitted in partial satisfaction of the

requirements for the degree Master of Arts

in Education

by

Dina Ali Naji Arch

Committee in charge:

Professor Karen Nylund-Gibson, Chair

Professor Rebeca Mireles-Rios

Professor Nancy Collins

June 2021

The thesis of Dina Ali Naji Arch is approved.

Dr. Rebeca Mireles-Rios

Dr. Nancy Collins

Dr. Karen Nylund-Gibson, Committee Chair

June 2021

ABSTRACT

Moderation with a latent class variable: An applied example

by

Dina Ali Naji Arch

Moderation analyses with a latent class variable allow researchers to study relations among exogenous (predictor or covariate) and endogenous (distal outcome) variables across the latent classes. Extending the latent class model to include auxiliary variables (both a predictor and an outcome) creates a context where the latent class variable serves as a moderating variable, referred to as a mixture regression. This thesis provides a pedagogical introduction on how to specify and interpret the moderation model with a latent class variable with the (ML) three-step manual approach (Asparouhov & Muthén, 2014) using data from the Longitudinal Survey of American Life (LSAL). Specifically, the latent class variable (science attitude classes) is hypothesized to moderate the relation between a predictor (science ability) and outcome (interest in science issues), controlling for the demographic variables. This thesis will outline the steps of the ML three-step approach, appropriate tables and visualizations used, and accompanying M*plus* syntax.

I. Introduction

Moderation analyses are a valuable tool in social science research and allow for a more nuanced understanding of the relationship between predictor and outcomes, specifically how the relations can be different across levels of the moderator. Traditionally, testing hypotheses with moderating variables are done using ordinary least squares (OLS) methods and introduced in the multiple regression context. First, an interaction term is created, and the significance of this interaction is used to assess moderation. In the presence of moderation, the main effects of the exogenous variable are estimated in a conditional fashion. For example, consider a binary moderator of an experimental group (treatment vs. control). In the presence of moderation, the relationship between the outcome and the predictor would depend on the experimental group’s level. In this example, for the treatment group, the relation between two variables could be positive and significant, but for the control group, it could be non-significant. More recently, this approach has been the focus of methodological development, extending the moderation model to answer the increasingly complex hypotheses using both observed (or measured) and unobserved (or latent) variables.

The most common method of incorporating moderators using latent variables is in a structural equation modeling (SEM) context. An advantage of using SEM for moderation analyses is to provide measurement error within the statistical model. Little et al. (2007) describe how to model contextual factors with moderators measured as continuous latent variables. They suggest an orthogonal approach to moderation with SEM by forming all possible products of the continuous indicators involved in each latent variable and removing any of the main-effect information. The new orthogonal indicators are then used as the latent interaction construct and then included in the SEM model. The SEM context easily accommodates continuous and categorical moderators. This paper aims to demonstrate the extension of the moderation model to include mixture models with a categorical latent variable, a latent class analysis (LCA) variable.

A. Moderation in the Mixture Modeling Context

Mixture modeling is a widely utilized statistical method in social science research used to identify unobserved subgroups within a population. Latent class analysis (LCA), a type of mixture model, uses a categorical or binary set of indicators to estimate the model and identify subgroups or "classes." The primary purpose of LCA is to find the number of classes based on the response patterns within the data. For a more comprehensive look at LCA, see Nylund-Gibson & Choi (2018).

In applied research, estimating latent class models to identify heterogeneity in the population is the first of several research questions. We can extend the latent class model to include auxiliary variables, including covariates, predictors, and distal outcomes, which provide a context to explore moderation. Moderation with a latent class variable allows the estimated regression parameters to be different across the latent classes (McLarnon et al., 2018). Additionally, moderation using latent class models may be useful to researchers interested in examining linear relationships among a predictor and an outcome across subsets of individuals. Finding evidence of moderation provides more information about the individuals in the latent classes, as opposed to a traditional regression where the moderators may not be person-centered. For example, Felix et al. (2019) used a latent class moderator to study how heterogeneity in flood exposure and its stressor on social-emotional health in youth. After identifying the latent classes of disaster exposure (high, moderate, community, and low exposure), they examined differences in the relations between life stressors (predictor) and social-emotional health (distal outcome) across the latent classes of disaster exposure.

Extending the latent class model to include auxiliary variables (both a predictor and an outcome) creates a context where the latent class variable serves as a moderating variable, referred to as a mixture regression. Specifically, we are moderating the relationship between exogenous (predictor or covariate) and endogenous (distal outcome) variables across the latent class, which allows for heterogeneity between variables that would otherwise be the same. While there are various approaches to including auxiliary variables into mixture models (e.g., one-step, classify-analyze, direct-inclusion), this paper will focus on one type, the maximum likelihood (ML) three-step manual approach (Asparouhov & Muthén, 2014). Mixture regression is a relatively new technique, and the subsequent section will detail the steps used in the analyses as well as the corresponding tables and figures required for each step.

B. Moderation using the ML Three-Step Method

The ML three-step approach is currently one of the recommended approaches for applying auxiliary variables into the latent class model. It has been shown to reduce parameter shifts in the model and be less biased than other approaches (Asparouhov & Muthén, 2014). The three modeling steps in the context of a mixture regression where the latent class variable is the moderator are as follows: 1) perform class enumeration of the unconditional LCA/LPA model, 2) determine measurement error of the modal class assignment, 3) specify the mixture regression with the effect of auxiliary variables are entered into the model. The three-step is described briefly here, highlighting how we would modify the specification for a moderation analysis, but the BCH method could be used.

1. Class Enumeration

In this first step, we decide how many classes should be used to represent the heterogeneity in the set of indicators. First, we start with identifying a one-class model and increasing the number of classes until a nominal increase in model fit or non-identification of the estimated model solution is found. Then, fit information from each model is tabulated and studied to decide on the number of classes. There are many considerations involved when selecting a latent class model and researchers use different criteria when doing so. Nylund, Asparouhov, and Muthén (2007) used simulation techniques to examine the performance of likelihood-based tests and relative model fit statistics, or information criteria (IC), to decide on the number of classes in LCA. Fit indices used in this paper that have been shown to identify the correct number of classes are as follows: Consistent Akaike's Information Criterion (CAIC), Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (aBIC), Approximate Weight of Evidence Criterion (AWE), Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT), and Bootstrapped Likelihood Ratio Test (BLRT). The researchers should create a table of all relevant fit indices and highlight the model suggested by each fit index. After selecting the appropriate latent class model, it may be necessary to re-estimate the latent class model and request additional information that will be used in step two. In M*plus*, it is essential to request the classification probabilities and modal class assignment variable (SAVEDATA: save= cprob;).

For the visual presentation of the latent class variable, it is recommended to consider the conditional item probability plot of the chosen solution. This plot is used to interpret and label the emergent latent classes as well as inspect the amount of endorsement for each indicator across each latent class and class size. The item probability plot is a key figure to include in the results section for the enumeration step.

2. Determine Measurement Error

After the enumeration step, the logits for the classification probabilities of the modal class assignment are extracted from the enumeration output. The logits are used in the final step to determine the measurement error of the modal class assignment. The modal class assignment is the class that an individual would be in if assigned one (e.g., the class with the highest posterior probability). Previously, the logits were entered manually into the syntax in step three. However, the Appendix provides information on how to extract these values from the M*plus* output and enter them into the syntax. For an example of how to specify a three-step model, see the code at the end of Nylund-Gibson et al. (2014).

3. Adding Auxiliary Variables

Finally, the new dataset created in the first step (which includes modal class assignment) and the logits extracted in the second step is ready to be used in the third and final step: specifying the moderation model with auxiliary variables. Specifically, a linear regression of the distal outcome(s) on the predictor(s) is freely estimated across each latent class. In M*plus*, this is done by repeating the regression in each of the class-specific statements (see Appendix).

C. Testing for Moderation

To test for moderation, or the equivalence of the slopes across the latent classes, we use the omnibus Wald chi-square test. If this test is significant, there is evidence that at least one slope is significantly different. Conceptually, a significant Wald test indicates that there is evidence that is statistically significant relations of the predictor and outcome across classes. To further evaluate which classes are significantly different, we conduct pairwise differences.

Additionally, we can test the equivalence of the regression intercepts, which in this context is the mean of the distal outcome, using the Wald chi-square test. For this test, we check for significant differences in the intercept across the latent classes. If there is evidence of a difference, then the pairwise comparisons of the intercepts are completed. A note when using M*plus*: the omnibus Wald tests must be conducted separately. This means the syntax in the third step would need to be run twice for two different Wald tests. The pairwise differences, however, can be entered into the syntax simultaneously. See Appendix for more details on how correctly identify the Wald tests.

Finally, while not necessarily evidence of moderation, we can look at individual regression slopes and intercepts across classes and report whether the slope for a specific class is significantly different than zero. Since there are several comparisons and model tests, we provided a table that supported the understanding of all the comparisons made (Table 1). Since we are estimating regression coefficients for each class, there are many ways to compare the estimated parameters. We can test the significance of the individual parameter (like we do with traditional regression models), the equivalence of the regression parameters across class, and pairwise differences.

To visually display evidence of moderation (significant differences in slopes) and mean differences (significant differences in intercepts), a table of slope and intercept values across latent classes and their significance, as well as a simple slopes graph and distal means bar chart, are recommended and will be illustrated in this paper using an applied example.

D. The Current Paper

While the mixture regression model is not new, it is underutilized in social science literature, likely because it can be hard to interpret the results. This paper provides a closer look into the modeling steps, interpretation, and visualization of a mixture regression model using an applied example using the Longitudinal Survey of American Life (LSAL) open-source dataset.

In the current study, we examined the cohorts at the twelfth-grade level to assess attitudes towards science issues related to STEM interest in science issues. Specifically, we are interested heterogeneity in twelfth-grade students' science attitudes and the relation between their science ability and their interest in science issues. This example is an extension of Ing & Nylund-

Gibson's (2013; 2017) work on early attitudes toward mathematics and science. In the recent paper (Ing & Nylund-Gibson, 2017), findings reveal that students have consistent attitudes towards math and science over time. In this paper, we hypothesize that, based on the science attitude class, the students' science ability will motivate how interested they are in science issues.

While there are papers using mixture regression, the amount of detail that authors can include in applied papers does not lend itself to helping new users understand how to apply the method in a new context. The purpose of this paper is to provide a walkthrough of a technique used for mixture regressions and the tables and figures that can be used to aid in interpretation. Specifically, an application of the maximum likelihood (ML) three-step method using auxiliary variables (Asparouhov & Muthén, 2014) is used with suggested tables and figures. We provide a walkthrough using this example in M*plus* (Muthén & Muthén, 1998-2021) to interpret results beyond what is usually presented in applied papers (see Appendix for M*plus* syntax).

II. Method

A. Sample

The dataset used in this example comes from the Longitudinal Study of American Life (LSAL; J.D Miller, 2010), funded by the National Science Foundation (NSF) in 1986. The LSAL was designed to study the development of student achievement in middle and high school and the relationship of those patterns to later career choices. The LSAL used telephone interviews and questionnaires to survey the two cohorts (younger and older) of seventh graders every year for seven years while they were in school (n = 5,945). The final sample consists of n = 2,488 students. The demographics in the final sample were predominantly white (74.2%), with an approximately equal number of females (49%) and males (51%). About 58% of the students' mothers have at least a high school diploma, ten percent had some college education, and 11% had a four-year college degree.

B. Measures

Using twelfth-grade student survey responses of the LSAL, we used student responses to five attitudinal variables about science to create the latent class variable. The latent class variable was used as a moderator to examine the relationship between science performance and interest in science issues. We then linked these classes to the covariates gender, ethnicity, underrepresented minority (See Table 2).

1. Science Attitudes

Students were asked how much they agree or disagree with five science issues. The rationale for selecting the science attitude items for the LCA is consistent with Ing and Nylund-Gibson (2017). The items chosen reflect a social cognitive career development perspective that highlights the students' self-efficacy, outcome expectancies, and personal goals. The questions all have the same response options on a Likert-type scale (strongly agree, agree, not sure, disagree, strongly disagree). Additionally, the items were dichotomized to stay consistent with previous analyses (Ing & Nylund-Gibson, 2017; Ing & Nylund-Gibson, 2013), where strongly agree and agree/not sure were coded as "1" and disagree/strongly disagree were coded as "0." The response options are used so that a "1" represents endorsing the science attitude, and a "0" represents that a student did not endorse that item.

2. Distal Outcome

Students responded to four items that asked about their interest in social issues. They rated their interest on a three-point scale: Not at all interested, moderately interested, and very interested. Four items related to science were chosen for analyses: Space exploration, scientific discoveries, inventions/technologies, and energy policy issues. The items were used to create a latent factor of students' interest in science issues. A factor, as opposed to a composite variable, was used as a distal outcome to reduce measurement error bias and provides the ability to look at the validity of the items using model fit indices. The overall goodness of fit indices in the final factor analysis model of the four items suggest the model fit the data well: $\chi 2(2) = 30.112$, $p = .00$, SRMR = .014, RMSEA = .063 (90% CI = .045 – .084), CFI = .99, TLI = .97. For identification in the moderation model, the factor means for one of the classes was fixed to zero, similar to what is done when exploring measurement invariance across groups (van de Schoot, Lugtig & Hox, 2012).

3. Predictors and Covariates

Science ability scores were used to predict the distal outcome of interest in science issues. Specifically, science IRT scores measured at the twelfth grade is used as a proxy for science ability. Gender, ethnicity, and socioeconomic status were used as covariates. Gender and ethnicity are treated as dichotomous variables. Ethnicity was dichotomized to represent students typically represented (Asian and White) and those typically underrepresented in STEM fields (African American, Hispanic, Native American). For ethnicity, or the underrepresented minority (URM) variable, "0" represents students who are typically represented in STEM fields (e.g., students who identify as White or Asian), and "1" denotes students who are typically underrepresented in STEM fields (e.g., African American, Hispanic, Native American). For gender, "0" represents males, and "1" represents females. Mother's educational attainment was used as a proxy for socioeconomic status (SES). The original variable in the LSAL dataset consisted of 9 levels: less than high school, high school graduation only, vocational or trade school, some college, associate degree, Bachelor's degree, Master's degree, Ph.D./MD, or 'I don't know'. The variable was collapsed for this study to include only five levels of education:

less than high school, high school diploma, some college, 4-year college, and an advanced degree.

4. Analysis Plan

The specification of moderation using a latent class variable is conducted in several steps, delineated here. The path diagram of the hypothesized model in Figure 1 is similar to moderation using linear regression, where the moderator (latent class variable) is hypothesized to moderate the relation between predictor (science ability scores) and the outcome (interest in science issues), controlling for the demographic variables.

First, class enumeration for the latent class variable using the five science attitude indicators (Table 3) is conducted using recommended approaches to enumeration without the auxiliary variables (Nylund-Gibson & Choi, 2018, Nylund-Gibson & Masyn, 2016). Multiple indicators of model fit are used to determine the final number of classes: Consistent Akaike's Information Criterion (CAIC), Bayesian Information Criterion (BIC), adjusted Bayesian Information Criterion (aBIC), Approximate Weight of Evidence Criterion (AWE), Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (VLMR-LRT), and Bootstrapped Likelihood Ratio Test (BLRT). Lower values of the BIC, CAIC, saBIC, and AWE indicate a better fitting model (Masyn, 2013). For the BLRT and VLMR-LRT, p values less than 0.05 indicate that the model has not significantly improved compared to the model with one less class (Nylund et al., 2007).

After identifying the best fitting latent class model, we specified the moderation model. We used the ML three-step specification (Asparouhov & Muthén, 2014; Vermunt, 2010) to estimate the latent class variable simultaneously and enable the relationship between the predictor and distal outcome to vary across the latent classes while holding the measurement of the latent class variable constant. The predictor variable, science ability, was scaled and centered. For comparison and identification, the factor mean for one of the classes, in this example, the third class (*Ambivalent with Minimal Utility*) was fixed to zero, similar to what is done in the context of factor mean comparison in measurement invariance. Alternatively, the researcher could swap out the mean of the factor to be zero of a different class if they are interested in other comparisons.

III. Results

The presentation of results begins with the descriptive statistics for all the variables used in the LCA and moderation model, including covariates. Following the descriptive statistics, the class enumeration results for the LCA models are presented, followed by the covariate and distal outcome results for the mixture moderation. Then, the mean slope differences across class and distal outcome means are visually presented to describe the mixture moderation model.

A. Descriptive Statistics

Table 2 presents the endorsement proportion for latent class indicators (attitudes in science) and the means and standard deviations for the covariates (ability and interest in science issues). Overall, all five items were highly endorsed by students, with the least endorsed item at 70% endorsement. The latent class indicator most endorsed by students was the "Science Helps Logical Thinking" item. The second most endorsed item was "Science is Useful in Everyday Problems," with 80% endorsement. Conversely, the least endorsed item is "I Enjoy Science," with 70% endorsement. Students may agree on the utility of science but may not enjoy the subject itself as much. The predictor of science ability (science IRT scores)'s average score is about 66, with the scores ranging from 27.01 and 99.30.

B. Latent Class Enumeration

A series of LCA models were fit, starting with a one-class model until non-convergence was achieved. Table 3 presents the fit indices used to compare the models with one through five latent classes. Three of the four information criteria (BIC, SABIC, and CAIC) reached a minimum value for the 4-class model. However, AWE values began increasing after the 3-class model, suggesting a 3-class model. The likelihood-based tests each pointed to the 4-class solution. The non-significant BLRT and VLMR – LRT p-value for the 5-class model shows support for the 4-class solution. Taken together, since five of the six fit indices suggest the 4-class model, the 4-class model was retained as the model to describe the heterogeneity of science attitudes in twelfth-grade students.

After choosing the 4-class model, the conditional item probability plot was used to interpret and label the four emergent latent classes (Figure 2). The classes were labeled *Pro-Science With Elevated Utility* (i.e., high probability of endorsement for all items), *Ambivalent with Elevated Utility* (i.e., moderate endorsement probability for the three items "Science is Useful," "Science Helps Logical Thinking," and "Need Science for a Good Job," with a higher probability of endorsement for the items "I Enjoy Science" and "Will Use Science Often as an Adult"), *Ambivalent with Minimal Utility* (i.e., moderate endorsement probability for the three items of "I Enjoy Science," "Science is Useful," and "Science Helps Logical Thinking," with a higher probability of endorsement for the items "Need Science for a Good Job" and "Will Use Science Often as an Adult"), and *Anti-Science with Minimal Utility* (i.e., low probability of endorsement for all items). The *Pro-Science with Elevated Utility* class makes up 64.65% of the sample, *Ambivalent with Elevated Utility* consisted of 10.59% of the sample, *Ambivalent with Minimal Utility* consisted of 13.58% of the sample, and *Anti-Science with Minimal Utility* made up 11.18% of the sample.

1. Covariates

The three covariates (URM, gender, and mother's education) were included in the model as predictors of the distal outcome. Table 4 presents the relations between the three covariates and the latent factor of interest in science issues. Underrepresented minority, gender, and mother's education are significant predictors of the distal latent outcome, interest in science issues, regardless of class.

C. Exploring the Moderating Effect of Science Attitudes

To interpret a mixture regression moderation, an overall test of equivalence of the regression of science issues on science ability across the latent classes was conducted using the omnibus Wald test. We first tested if there was a relation between the slopes across the latent classes. If there was evidence of a relation (e.g., significant Wald test), pairwise comparisons of the regression slopes across classes are warranted. Next, a second omnibus test was used to explore differences in the distal outcome means across classes. If significant, pairwise distal outcome comparisons across classes are warranted. Figure 3 presents the relationships between the slope and intercept parameters estimated independently for each latent class. In this graph, the x-axis is fixed to be uncentered and unscaled. Additionally, the order of presentation corresponds to Table 1. Table 5 presents an organized table of the slope and intercept value across science attitude classes.

1. Slope Differences

In this example, there is evidence of a significant moderation ($\chi^2(3) = 15.356$, $p < .001$). That is, there is a statistically different relations between the predictor (science scores) and the distal outcome (interest in science issues) across at least one of the classes. Specifically, there are differences in the relation between science ability and interest in science issues between at least one pair of science attitude classes.

a. Pairwise slope differences

To further investigate which class-specific relations differ, pairwise comparisons of the regression slopes on the distal outcome, issues in science, were carried out. When examining the pairwise slope differences for issues in science regressed on science ability, the *Pro-Science with Elevated Utility* class was significantly different from the *Ambivalent with Elevated Utility* and *Ambivalent with Minimal Utility* class, $p < .05$. Specifically, the rate at which science ability predicts interest in science issues differs among these classes. Figure 3 visually presents the relations between science ability and interest in science issues across each class. There were no other significant slope differences across classes.

b. Regression coefficients

Additionally, each regression between the predictor and outcome was examined across classes. The regressions in the *Pro-Science with Elevated Utility* and *Ambivalent with Minimal Utility* class were significantly different from zero, *b*s=0.097 and 0.173, respectively. This implies that for students in the *Pro-Science with Elevated Utility* and *Ambivalent with Minimal Utility* class, their interest in science issues significantly increases as their science performance increases. For students in the remaining two classes, *Ambivalent with Elevated Utility* and *Anti-Science with Minimal Utility*, the regression for *issues in science* on *science ability* were non-significant. This is one form of moderation in this context because the relationship between students' science scores and their interest in science differs depending on the students' attitudes towards science class.

2. Distal Outcome Differences

There was evidence that there are significant differences in the distal outcome means across the science attitude classes, $\chi^2(1) = 7.598$, $p < .001$. This suggests that at least one pairwise difference of the distal outcomes means is significant across classes. Figure 4 presents the means of the science ability (Science IRT scores; grand mean-centered and scaled) across the classes of Science Attitudes.

a. Pairwise distal outcome differences

To further investigate which class-specific relations differ, pairwise comparisons of the distal outcome, issues in science, across each class were studied. Pairwise tests found significant differences between *Pro-Science with Elevated Utility* class and *Ambivalent with Elevated Utility* class, p < .001, as well as *Anti-Science with Minimal Utility* class, p < .001. Additionally, there was a statistically significant difference between the *Ambivalent with Elevated Utility* class and the *Anti-Science with Minimal Utility* class, p < .05. This implies that the two classes, on average, have significantly different interests in science issues. Specifically, those in the *Pro-Science with Elevated Utility* class have more science interests than those in the *Anti-Science with Minimal Utility* and *Ambivalent with Elevated Utility* classes. Additionally, those in the *Ambivalent with Elevated Utility* class have more science interest than those in the *Anti-Science with Minimal Utility* class. There was no other significant distal outcome mean differences across classes.

b. Intercept coefficients

The mean of the distal outcome factor, interest in science issues, was set to zero for the *Ambivalent with Minimal Utility* for measurement identification when adding the latent variable. This class was used as the reference class. Thus, the mean of the factor is set to zero, and others are compared to it. Compared to the Ambivalent w/ Minimal Utility Value class, students in the *Pro-Science with Elevated Utility* class had increased science ability (M = 0.176), they showed increased science ability. In comparison, students in the *Anti-Science with Minimal Utility* class performed lower on science scores (M = -0.143), on average.

IV. Discussion

This paper provided a pedagogical example of mixture regression analysis using a latent class moderator. Traditionally, testing hypotheses with moderating variables are commonly done using multiple regression and SEM models. This paper aimed to extend the moderation model to include categorical latent class variables using an applied example. Specifically, we moderated the relationship between an exogenous (predictor or covariate) and endogenous (distal outcome) latent variable among the latent class moderator. Additionally, to support this understanding, we provide a walkthrough of how to specify a mixture regression model using the ML 3-step method in M*plus* with annotated output (see Appendix), providing tables and figures that can be used to understand the moderation.

The applied example used data from the Longitudinal Study of American Life (LSAL) to walk through the moderation steps using ML 3-step and its interpretation. There was evidence of moderation utilizing a series of model testing steps, which was highlighted. That is, there is a difference in relations between science ability and interest in science issues across the latent classes of science attitudes. Since there are many class-specific parameters of the regression, there are multiple ways to test the equivalence of these parameters. We demonstrated these comparisons using the LSAL example. Specifically, the example looked at heterogeneity in twelfth-grade students' science attitudes and the relation between their science ability and their interest in science issues. We found that the relationship between science scores and interest in science issues depends on the classes of twelfth-grade science attitudes. Following these tests, pairwise comparisons of the slopes and intercepts were completed.

Overall, students in all classes had positive slopes, indicating a positive relationship between science scores and their interest in science issues. Those in the *Pro-Science with Elevated Utility* class had statistically significant slope differences than those in the *Ambivalent with Elevated Utility* class and *Ambivalent with Minimal Utility* class. Finally, and unsurprisingly, those in the *Pro-Science with Elevated Utility* class had a higher interest in science issues than those in the *Ambivalent with Elevated Utility* class and *Anti-Science with Minimal Utility* class.

Some limitations to this study include the way the distal outcome was identified. The *Ambivalent with Minimal Utility* class could not be compared as it was set to zero for measurement identification. Additionally, other methods can be used to conduct mixture regression analyses, such as the BCH approach (Vermunt, 2010). This paper only examined the specification, interpretation, and visualization of the ML 3-step. This paper also provides an example using one predictor and one distal outcome.

Moderation in the mixture regression context enables researchers to study complex relationships, as shown in this applied example. In this example, we only include one predictor and one outcome, but more auxiliary variables could be included. Future pedagogical examples may consist of complex models with additional auxiliary variables in moderation models with LCA as the moderator. This would be interpreted as interaction within a class that varies across classes. Additionally, future examples could focus on different class-specific models. Finally, this example used the ML 3-step for estimation. Other methods can be used, such as the BCH (Vermunt, 2010).

References

1. Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using M*plus*. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(3), 329-341. https://doi.org/10.1080/10705511.2014.915181

2. Felix, E. D., Binmoeller, C., Nylund-Gibson, K., Benight, C. C., Benner, A. D., & Terzieva, A. (2019). Addressing disaster exposure measurement issues with latent class analysis. *Journal of Traumatic Stress*, *32*(1), 56–66. https://doi.org/10.1002/jts.22368

3. Ing, M., & Nylund-Gibson, K. (2013). Linking early science and mathematics attitudes to long-term science, technology, engineering and mathematics career attainment: Latent class analysis with proximal and distal outcomes. *Educational Research and Evaluation*, *19*(6), 510–524.

4. Ing, M., &; Nylund-Gibson, K. (2017). The importance of early attitudes towards mathematics and science. Teachers College Record, *119*(5), 1 -31.

5. Little, T. D., Bovaird, J. A., Card, N. A., Preacher, K. J., & Crandall, C. S. (2007). Structural equation modeling of mediation and moderation with contextual factors. *Modeling Contextual Effects in Longitudinal Studies* (pp. 207–230). Lawrence Erlbaum Associates.

6. Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), *The Oxford handbook of quantitative methods* (Vol. 2: Statistical analysis, pp. 551–611). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199934898.013.0025

7. McLarnon, M. J., & O'Neill, T. A. (2018). Extensions of auxiliary variable approaches for the investigation of mediation, moderation, and conditional effects in mixture models. Organizational Research Methods, 21(*4*), 955-982.

8. Miller, J. D. (2010). *Longitudinal Study of American Youth, 1987-1994, and 2007*. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

9. Muthén, L.K. and Muthén, B.O. (1998-2021). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.

10. Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

11. Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Scienc*e, *4*(4), 440. https://doi.org/10.1037/tps0000176

12. Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A Latent Transition Mixture Model Using the Three-Step Specification. *Structural Equation Modeling: A*

Multidisciplinary Journal, *21*(3), 439–454. https://doi.org/10.1080/10705511.2014.915375

13. Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling*, *23*(6), 782–797. https://doi.org/10.1080/10705511.2016.1221313

14. R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org

15. van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486–492. https://doi.org/10.1080/17405629.2012.686740

16. Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 450-469.https://doi.org/10.1093/pan/mpq025

V. Tables and Figures

Table 1

Stages of Analyses for Slope and Intercept Estimation

Test	Stage	Test	M*plus* Syntax	Result
Slope	I	Wald Test: Is the relation (slope) different across the latent classes? $\beta_{11} = \beta_{12} = \ldots = \beta_{1c}$	MODEL TEST: !Wald Test for Slopes B11=B12; B12=B13;	If significant, we have evidence of moderation and interaction. Proceed to stage II to test for pairwise differences

	II	Pairwise differences: Where are the slope differences?
		Is $\beta_{11} = \beta_{12}$?
		Is $\beta_{11} = \beta_{12}$?
			MODEL CONSTRAINT:
			new (slope12, slope13, slope23);
			slope12=b11-b12;
			slope13=b01-b13;
			slope23=b12-b13;
				If significant, there is a slope difference between the two classes examined.

	III	Regression coefficient: Is the slope for a specific class significantly different than zero? $\beta_{1c} = 0$?	Already exists in the %OVERALL% statement (B11). Ex: %C#1% [DISTAL] (B01); DISTAL; DISTAL on PREDICTOR(B11);	If significant, the slope in the class examined is significantly different from zero
Intercept	I	Wald Test: Are the distal means (intercept) different across the latent classes? $\beta_{01} = \beta_{02} = \ldots = \beta_{0c}$	MODEL TEST: !Wald Test for Intercepts B01=B02; B02=B03;	If significant, we have evidence of distal mean differences. Proceed to stage II to test for pairwise differences

	II	Pairwise differences: Where are the pairwise differences?
		Is $\beta_{01} = \beta_{02}$?
		Is $\beta_{01} = \beta_{02}$?
			MODEL CONSTRAINT:
			new (int12, int13, int23);
			int12=b01-b02;
			int13=b01-b03;
			int23=b02-b03;
				If significant, there is an intercept difference between the two classes examined.

	III	Intercept coefficient: Is the intercept for a specific class significantly different than zero? $\beta_{0c} = 0$?	Already exists in the %OVERALL% statement (B01). Ex: %C#1% [DISTAL] (B01); DISTAL; DISTAL on PREDICTOR(B11);	If significant, the intercept in the class examined is significantly different from zero

Table 2

Descriptive Statistics for Latent Class Indicators and Auxiliary Variables

Item Label	Endorsement Proportion or Mean (SD)	*n*
Latent Class Indicators		
I Enjoy Science	.70	2351
Science is Useful in Everyday Problems	.80	2670
Science Helps Logical Thinking	.85	2840
Need Science for a Good Job	.73	2439
Will Use Science Often as an Adult	.77	2573
Predictor		
Science IRT Score*	65.846 (11.65)	2826
Distal Outcome (Interest in Science Issues)		
Space Exploration	1.89 (0.68)	2471
Science Issues	2.00 (0.69)	2458
New Technologies	2.06 (0.67)	2462
Energy Policy Issues	1.77 (0.67)	2468

* Uncentered mean and standard deviation.

Table 3

Fit Statistics for Class Enumeration For 12th Grade Science Attitudes

Model	*K*	LL	BIC	SABIC	CAIC	AWE	BLRT *p*	VLMR LRT *p*
n =3364	1	-8884.78	17810.15	17794.27	17815.15	17865.76	-	-
	2	-7164.05	14417.43	14382.48	14428.43	14539.76	<.001	<.001
	3	-7054.07	14246.20	14192.18	14263.20	**14435.25**	<.001	<.001
	4	-7013.93	**14214.65**	**14141.56**	**14237.65**	14470.43	<.001	<.001
	5	-7011.08	14257.66	14165.51	14286.66	14580.16	**.667**	**.469**

Note. K = number of classes; LL = model log likelihood; BIC = Bayesian information criterion; SABIC = sample size adjusted BIC; CAIC = consistent Akaike information criterion; AWE = approximate weight of evidence criterion; BLRT = bootstrapped likelihood ratio test; VLMR-LRT = Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test; *p* = *p*-value; **Bold** = best fit statistic for each individual statistic.

Table 4

Relations Between the Covariates and the Distal Outcome

Covariate	*Estimate*
Underrepresented	.050*
Female	-.156***
Mother's Education	.077*

Note. *Anti-science w/ Minimal Utility Value* as the reference class. *p < .05. ***p < .001

Table 5

Slope and Intercept Values Across Science Attitudes Classes

		Science Ability (X)		→	Interest in Science Issues (Y)	
		(*Slope*)			(*Intercept*)	
k	Class Label	Estimate (*se*)	Sig. Class Differences?		Estimate (*se*)	Sig. Class Differences?
C1 *n* = 208	*Ambivalent w/ Elevated Utility Value*	0.077 (.039) *	C3		0.005 (.052)	C3, C4
C2 *n* = 297	*Ambivalent w/ Minimal Utility Value*	0.097 (.035) **	C3		0†	C3, C4
C3 *n* = 1702	*Pro-Science w/Elevated Utility Value*	0.173 (.012) **	C1, C2		0.176 (.033) **	C1, C2, C4
C4 *n* = 281	*Anti-science w/ Minimal Utility Value*	0.080 (.051)	None		-0.143 (.051) **	C1, C2, C3

Note. The Science Ability column represents slope values, and the Interest in Science Issues column represents intercept or mean values, $*p < .05$. $**p < .01$

† Intercept and *se* for C2 are not estimated because the mean is fixed to zero for identification of the latent factor.

Figure 1

Path Diagram: The Relationship Between Science Ability (Science IRT scores) and Issues in Science Moderated by Latent Class Attitudes Variable

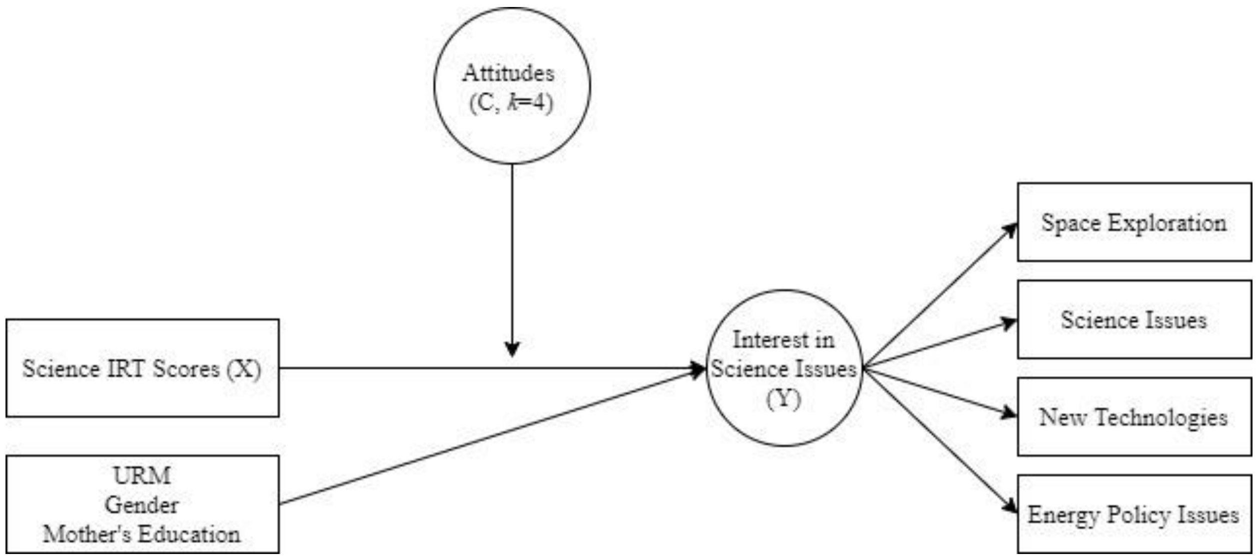


Figure 2

Proportions of Twelfth Graders Endorsing Each Item by Attitudinal Class

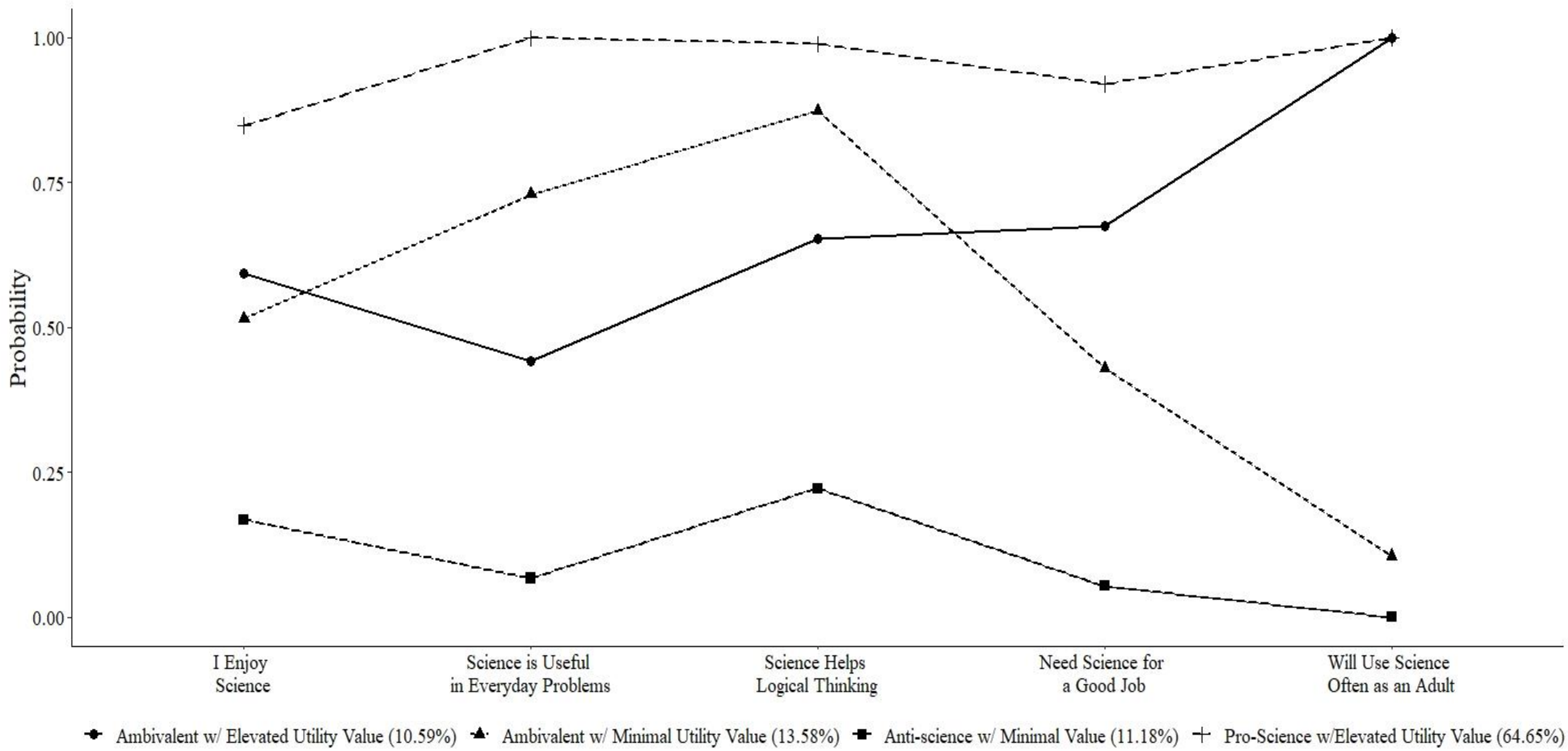


Figure 3

Simple Slopes Graph of Science Ability (Science IRT scores) Predicted by Issues in Science

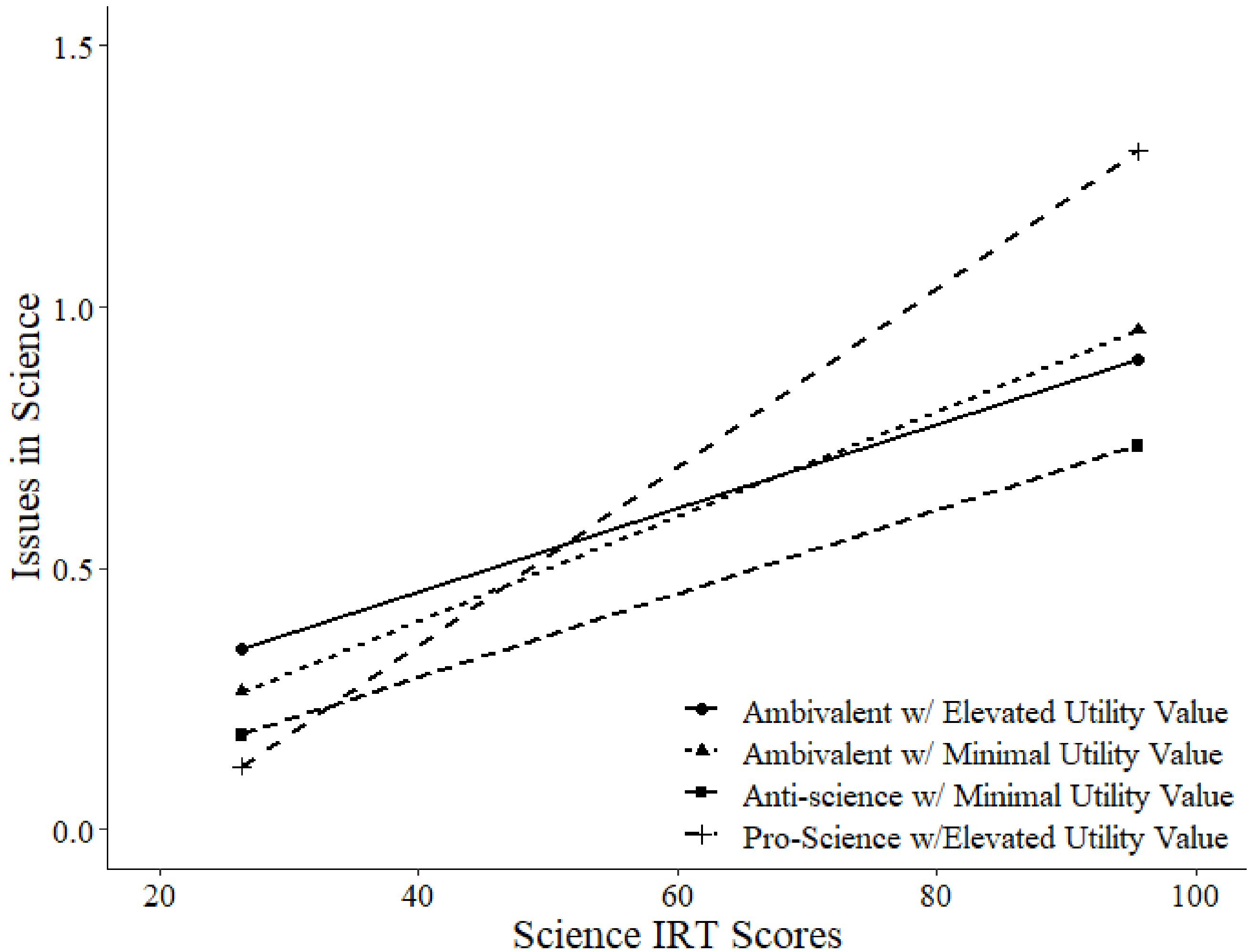
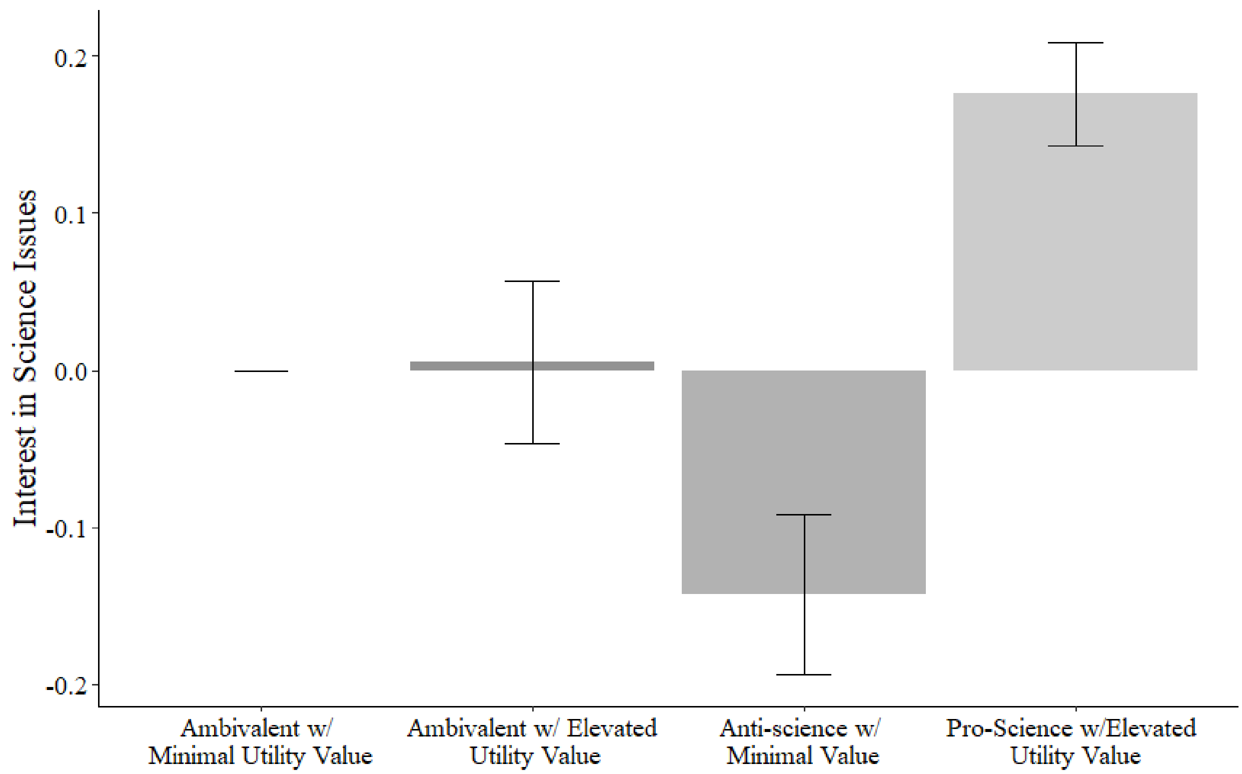


Figure 4

Means of Interest in Science Issues Across Classes of Science Attitudes



VI. Appendix

Annotated Output of the Manual ML Three-Step in M*plus*

This Appendix walks through the output of three different M*plus* runs using the Longitudinal Study of American Life (LSAL) example: descriptive statistics, class enumeration, and moderation model. Only relevant output is shown that corresponds to what is discussed in the paper. Below is a table with variable names and their description. *Note:* Comments in brown are notes and not part of the syntax. Notes can be included in M*plus* using exclamation points.

	Variable Name	**Description**
Latent Class Indicator	KA47A	I Enjoy Science
	KA47H	Science is Useful in Everyday Problems
	KA47I	Science Helps Logical Thinking
	KA47K	Need Science for a Good Job
	KA47L	Will Use Science Often as an Adult
Predictor	KSCIIRT	Science IRT Score
Distal Outcome	KA9B	Space Exploration
	KA9D	Science Issues
	KA9G	New Technologies
	KA9K	Energy Policy Issues
Covariates	URM	Under-represented Minority (0 = represented, 1 = under-represented)
	FEMALE	Sex (0 = male, 1 = female)
	MOTHED	Mother's Education (0 = less than high school, 1 = high school diploma, 2 = some college, 3 = 4-year college, 4 = an advanced degree)

Descriptive Statistics

Input Syntax

Below is input syntax to call descriptive statistics.

```
  TITLE: LSAL Descriptive Statistics;

  DATA:
  FILE = "LSAL_data.dat";

  VARIABLE:
    NAMES = KA46A KA46H KA46I KA46K KA46L KA47A KA47H KA47I KA47K
KA47L URM FEMALE
       MOTHED KSCIIRT KMTHIRT KA9B KA9D KA9G KA9K; !Column Names in
order
    MISSING=.; ! Identify missing value
    USEVAR = URM FEMALE MOTHED KSCIIRT KMTHIRT KA9B KA9D KA9G KA9K
      KA47A KA47H KA47I KA47K KA47L; ! Select variables to examine
    CATEGORICAL = KA47A KA47H KA47I KA47K KA47L URM FEMALE MOTHED;
! Identify the categorical variables

  DEFINE:
        KSCIIRT = KSCIIRT/10; ! In this example, we scale and
center our predictor
      CENTER KSCIIRT (GRANDMEAN);

  ANALYSIS:
        TYPE=basic; ! Identified for basic analysis (descriptive
statistics)

  OUTPUT:
        sampstat; ! Provides descriptive statistics
```

Annotated Output

Sample Statistics

The last part of the output (*UNIVARIATE SAMPLE STATISTICS*), is what we can evaluate and use in our descriptive statistics table for the continuous variables (Table 2).

```
UNIVARIATE SAMPLE STATISTICS

     UNIVARIATE HIGHER-ORDER MOMENT DESCRIPTIVE STATISTICS

         Variable/       Mean/      Skewness/   Minimum/ % with
Percentiles
        Sample Size    Variance    Kurtosis    Maximum  Min/Max
20%/60%     40%/80%     Median
```

```
     KSCIIRT                0.000       -0.351       -3.947     0.04%
-0.902      -0.219      0.040
                 2826.000     1.357        0.023        2.971     0.04%
0.301       1.000
     KMTHIRT               68.326       -0.289       27.010     0.04%
56.910      65.470     68.910
                 2780.000   186.300       -0.316       99.300     0.18%
72.520      80.120
     KA9B                   1.855        0.186        1.000    31.32%
1.000       2.000       2.000
                 3487.000     0.461       -0.847        3.000    16.86%
2.000       2.000
     KA9D                   1.963        0.050        1.000    25.99%
1.000       2.000       2.000
                 3470.000     0.481       -0.923        3.000    22.28%
2.000       3.000
     KA9G                   2.026       -0.033        1.000    22.09%
1.000       2.000       2.000
                 3476.000     0.468       -0.862        3.000    24.74%
2.000       3.00
     KA9K                     1.767        0.298        1.000    36.33%
1.000       2.000       2.000
                 3476.000       0.440       -0.782        3.000    13.06%
2.000       2.000
```

Proportion and Counts

Earlier in the output (*`UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES`*), is what we can evaluate and use in our descriptive statistics table for the categorical variables (Table 2).

```
UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES

    URM
      Category 1    0.777        4313.000
      Category 2    0.223        1241.000
    FEMALE
      Category 1    0.509        3026.000
      Category 2    0.491        2919.000
    MOTHED
      Category 1    0.147         855.000
      Category 2    0.580        3365.000
```

```
      Category 3    0.103           598.000
      Category 4    0.118           684.000
      Category 5    0.052           303.000
    KA47A
      Category 1    0.300          1006.000
      Category 2    0.700          2351.000
    KA47H
      Category 1    0.200           669.000
      Category 2    0.800          2670.000
    KA47I
      Category 1    0.148           492.000
      Category 2    0.852          2840.000
    KA47K
      Category 1    0.270           900.000
      Category 2    0.730          2439.000
    KA47L
      Category 1    0.233           782.000
      Category 2    0.767          2573.000
```

Note. The sample sizes presented in Table 2 are taken from later outputs to account for missingness in the analyses.

Moderation using the ML Three-Step Method

1. Class Enumeration

In the first step of the ML three-step, we decide how many classes should represent the heterogeneity in the set of indicators. First, we start with identifying a one-class model and increasing the number of classes until a nominal increase in model fit or non-identification of the estimated model solution is found. See Nylund-Gibson & Choi (2018) for a comprehensive review on enumeration methods. Below is the syntax for the four-class model. Change the estimation of class by replacing the 4 in `CLASSES = c(4);` and re-run the model in M*plus*.

An important note: Under *SAVEDATA:*, the classification probabilities and modal class assignment are requested to be saved into a new dataset. This is not necessary to enter into the syntax until after the latent class model is selected.

Input Syntax

```
TITLE: LSAL 4-Class Model;
DATA:
FILE = "LSAL_data.dat";
```

```
  VARIABLE:
    NAMES = KA46A KA46H KA46I KA46K KA46L KA47A KA47H KA47I KA47K
KA47L URM FEMALE
       MOTHED KSCIIRT KMTHIRT KA9B KA9D KA9G KA9K;
    MISSING=.;
    USEVAR = KA47A KA47H KA47I KA47K KA47L;
    CATEGORICAL = KA47A KA47H KA47I KA47K KA47L; ! Identified as
categorical for binary LCA
    CLASSES = c(4); ! Class 4
    AUXILIARY = URM FEMALE MOTHED KSCIIRT KA9B KA9D KA9G KA9K; !
Identifying auxiliary variables

  ANALYSIS:
    ESTIMATOR = mlr;
    TYPE = mixture;
    STARTS = 500 100; ! Starting values
    OPTSEED = 364676; ! set seed to replicate analyses at the same
log-likelihood and initial starts
  !SAVEDATA: ! Only keep this when re-running the chosen latent
class model
    !FILE = savedata.dat;
    !SAVE = cprob;

  PLOT:
    TYPE = plot3;
    SERIES = KA47A KA47H KA47I KA47K KA47L(*);
```

Annotated Output

Sample Size

At the beginning of an LCA output, we can see our sample size and the number of dependent and categorical variables used. We are estimating one categorical variable (latent class variable) and five indicator variables.

```
SUMMARY OF ANALYSIS

Number of groups                                                1
Number of observations                                       3364

Number of dependent variables                                   5
Number of independent variables                                 0
```

```
Number of continuous latent variables                           0
Number of categorical latent variables                          1
```

Proportion and Counts

Here, we can see the proportions and counts for each indicator variable. *`Category 1`* is no endorsement and, *`Category 2`* is the endorsement of the indicator variables.

```
UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES

    KA47A
      Category 1    0.300         1006.000
      Category 2    0.700         2351.000
    KA47H
      Category 1    0.200          669.000
      Category 2    0.800         2670.000
    KA47I
      Category 1    0.148          492.000
      Category 2    0.852         2840.000
    KA47K
      Category 1    0.270          900.000
      Category 2    0.730         2439.000
    KA47L
      Category 1    0.233          782.000
      Category 2    0.767         2573.000
```

Class Size

Here, we can find class sizes. For example, 10.588% of the sample are in Class 1. *Important note*: each time the model is re-run, there is a chance of the classes rearranging. Always check the class sizes and probabilities (shown next) when referring to the classes. Use *OPTSEED* (See Mplus manual) in the input syntax to set the seed for analysis and avoid class rearrangement.

```
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

    Latent
   Classes

      1        356.16452          0.10588
      2        376.19409          0.11183
```

```
         3          456.68905          0.13576
         4         2174.95234          0.64654
```

Labels of Latent Class Based on M*plus* Output

Latent Class	**Label**
1	Ambivalent with Elevated Utility Value
2	Anti-Science with Minimal Utility Value
3	Ambivalent with Minimal Utility Value
4	Pro-Science with Elevated Utility Value

Conditional Item Probabilities

Below are is the output that identifies the conditional item probabilities. The values under *Estimate* are the conditional item probabilities for each indicator variable across each latent class. Recall that `Category 1` is no endorsement and `Category 2` is the endorsement of the indicator variables. For example, the probability of those in *Class 1* endorsing item *KA47A* is 0.593. The endorsement of the conditional item probabilities should be plotted to visualize the latent class variable.

```
RESULTS IN PROBABILITY SCALE

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Latent Class 1

 KA47A
    Category 1         0.407      0.042      9.619      0.000
    Category 2         0.593      0.042     14.020      0.000
 KA47H
    Category 1         0.558      0.094      5.910      0.000
    Category 2         0.442      0.094      4.682      0.000
 KA47I
    Category 1         0.346      0.069      5.034      0.000
    Category 2         0.654      0.069      9.499      0.000
 KA47K
    Category 1         0.326      0.045      7.217      0.000
    Category 2         0.674      0.045     14.919      0.000
 KA47L
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
```

```
Latent Class 2

 KA47A
    Category 1         0.832        0.027       30.935        0.000
    Category 2         0.168        0.027        6.253        0.000
 KA47H
    Category 1         0.932        0.027       33.919        0.000
    Category 2         0.068        0.027        2.459        0.014
 KA47I
    Category 1         0.778        0.038       20.632        0.000
    Category 2         0.222        0.038        5.884        0.000
 KA47K
    Category 1         0.946        0.017       55.687        0.000
    Category 2         0.054        0.017        3.176        0.001
 KA47L
    Category 1         1.000        0.000        0.000        1.000
    Category 2         0.000        0.000        0.000        1.000

Latent Class 3

 KA47A
    Category 1         0.485        0.028       17.157        0.000
    Category 2         0.515        0.028       18.239        0.000
 KA47H
    Category 1         0.271        0.038        7.185        0.000
    Category 2         0.729        0.038       19.361        0.000
 KA47I
    Category 1         0.126        0.029        4.415        0.000
    Category 2         0.874        0.029       30.498        0.000
 KA47K
    Category 1         0.571        0.032       18.014        0.000
    Category 2         0.429        0.032       13.542        0.000
 KA47L
    Category 1         0.895        0.153        5.862        0.000
    Category 2         0.105        0.153        0.690        0.490

Latent Class 4

 KA47A
    Category 1         0.151        0.010       15.652        0.000
    Category 2         0.849        0.010       87.684        0.000
 KA47H
    Category 1         0.001        0.011        0.072        0.943
    Category 2         0.999        0.011       87.764        0.000
```

```
 KA47I
    Category 1          0.011        0.007        1.633        0.102
    Category 2          0.989        0.007      147.300        0.000
 KA47K
    Category 1          0.080        0.009        8.712        0.000
    Category 2          0.920        0.009       99.962        0.000
 KA47L
    Category 1          0.000        0.000        0.000        1.000
    Category 2          1.000        0.000        0.000        1.000
```

2. Determine Measurement Error

After the enumeration step, the logits for the classification probabilities of the modal class assignment are extracted from the output created in the enumeration step. These logits are used in the third and final step to determine the measurement error of the modal class assignment. There are no models estimated in this step, only the extraction of the logits to be used in the final step.

Annotated Output

Logits for Classification Probabilities

Below is appended output from the enumeration step.

```
CLASSIFICATION QUALITY

Logits for the Classification Probabilities for the Most Likely
Latent Class Membership (Column)
by Latent Class (Row)

              1          2          3          4

    1       0.874     -6.329     -5.617      0.000
    2      -1.487      6.772      4.393      0.000
    3      -0.606      0.392      2.431      0.000
    4      -4.425    -13.804    -10.500      0.000
```

The logits presented are entered manually into the syntax in step three. See the next step on how these logits are included in the syntax.

3. Adding Auxiliary Variables

Finally, the new dataset created in the first step (which includes modal class assignment) and the logits extracted in the second step is ready to be used in the third and final step: specifying the moderation model with auxiliary variables. Additionally, we can test the equivalence of the regression intercepts, which in this context is the mean of the distal outcome, using the Wald chi-square test. Moderation is thought to occur when at least one slope is different, as evidenced by a significant Wald chi-square test. However, the omnibus Wald tests must be conducted separately.

Input Syntax

A linear regression of the distal outcome(s) on the predictor(s) is freely estimated across each latent class to test for moderation. In M*plus*, this is done by repeating the regression in each of the class-specific statements. See Figure 1 for the path diagram that corresponds with this syntax.

```
  TITLE: LSAL Moderation;
  DATA:
    FILE = "savedata.dat";

  VARIABLE:
    NAMES = KA47A KA47H KA47I KA47K KA47L URM FEMALE MOTHED KSCIIRT
KA9B KA9D KA9G KA9K
       BCHW1 BCHW2 BCHW3 BCHW4 CPROB1 CPROB2 CPROB3 CPROB4 N;
    MISSING=.;
    USEVAR = URM FEMALE MOTHED KSCIIRT KA9B KA9D KA9G KA9K N;
    CLASSES = c(4);
    NOMINAL = N; ! N is the modal class assignment from the dataset
we created in step 1

  DEFINE:
    KSCIIRT = KSCIIRT/10; ! Scale the predictor
    CENTER KSCIIRT (GRANDMEAN); ! Center the predictor

  ANALYSIS:
    ESTIMATOR = mlr;
    TYPE = mixture;
    STARTS = 0;
    ITERATIONS = 1000;

  MODEL:
    !Covariates: URM FEMALE MOTHED KSCIIRT
    !Distal: ISSUES
```

```
    %OVERALL%
    ISSUES by KA9B KA9D KA9G KA9K; ! Creating the factor for the
distal outcome

    ISSUES on URM FEMALE MOTHED; ! Covariates -> Issues in Science
    ISSUES on KSCIIRT; ! Science Scores -> Science Issues

           %C#1% ! Class 1
  [N#1@0.874]; ! The modal class assignment variable (N) and logits
are entered here to specify measurement error
  [N#2@-6.329];
  [N#3@-5.617];
        [ISSUES] (B01); ! Estimation of intercept
        ISSUES;
        ISSUES on KSCIIRT(B11); ! Estimation of slope (Science
Scores -> Science Issues)

          %C#2% ! Class 2
  [N#1@-1.487];
  [N#2@6.772];
  [N#3@4.393];
        [ISSUES] (B02);
        ISSUES;
        ISSUES on KSCIIRT(B12);

          %C#3% ! Class 3
  [N#1@-0.606];
  [N#2@0.392];
  [N#3@2.431];
        [ISSUES@0] (B03); ! Here, we set a class equal to zero for
measurement identification of the latent factor
        ISSUES;
        ISSUES on KSCIIRT(B13);

          %C#4% ! Class 4
  [N#1@-4.425];
  [N#2@-13.804];
  [N#3@-10.5];
        [ISSUES] (B04);
        ISSUES;
        ISSUES on KSCIIRT(B14);

  MODEL TEST:
      !Omnibus test 1
```

```
!Only one omnibus test may be estimate at one time, the second one
is commented out here. After estimating this first omnibus test of
slopes, the second omnibus test of intercept may be estimated after
removing the "!" the second test and commenting out the first test.
       B11=B12;
       B12=B13;
       B13=B14;
      !Omnibus test 2
       !B01=B02;
       !B02=B03; ! Because we set class three equal to zero, we
cannot include its intercepts in the omnibus test
       !B03=B04;

  MODEL CONSTRAINT: ! Pairwise differences for slope and intercepts
can be tested simultaneously
      new (slope12, slope13, slope14, slope23, slope24, slope34,
           int12, int14, int24);
      slope12=B11-B12;
      slope13=B11-B13;
      slope14=B11-B14;
      slope23=B12-B13;
      slope24=B12-B14;
      slope34=B13-B14;
      int12=B01-B02; ! Class three not included
      int14=B01-B04;
      int24=B02-B04;
```

Annotated Output

Sample Statistics

Presented are the updated sample statistics accounting for listwise deletion in the analyses.

```
UNIVARIATE SAMPLE STATISTICS

     UNIVARIATE HIGHER-ORDER MOMENT DESCRIPTIVE STATISTICS

         Variable/         Mean/     Skewness/   Minimum/ % with
Percentiles
        Sample Size      Variance    Kurtosis    Maximum   Min/Max
20%/60%     40%/80%    Median

     KA9B                      1.894       0.132       1.000    28.81%
```

```
1.000       2.000       2.000
             2471.000       0.459       -0.833       3.000   18.21%
2.000       2.000
     KA9D                   2.003       -0.004       1.000   23.39%
1.000       2.000       2.000
             2458.000       0.471       -0.876       3.000   23.68%
2.000       3.000
     KA9G                   2.059       -0.070       1.000   19.86%
2.000       2.000       2.000
             2462.000       0.453       -0.795       3.000   25.75%
2.000       3.000
     KA9K                   1.773        0.283       1.000   35.70%
1.000       2.000       2.000
             2468.000       0.435       -0.765       3.000   12.97%
2.000       2.000
     URM                    0.204        1.471       0.000   79.62%
0.000       0.000       0.000
             2488.000       0.162        0.163       1.000   20.38%
0.000       1.000
     FEMALE                 0.521       -0.084       0.000   47.91%
0.000       0.000       1.000
             2488.000       0.250       -1.993       1.000   52.09%
1.000       1.000
     MOTHED                 2.385        1.099       1.000   11.09%
2.000       2.000       2.000
             2488.000       1.004        0.584       5.000    5.02%
2.000       3.000
     KSCIIRT                0.000       -0.368      -3.978    0.04%
-0.866      -0.215       0.032
             2488.000       1.338        0.109       2.940    0.04%
0.294       0.986
```

Slope Differences

Below is the first omnibus Wald test results for slope differences. In this example, this is evidence of a significant moderation because of the significant Wald test. That is, there is a significant relationship between the predictor (science scores) and the distal outcome (interest in science issues) across at least one of the classes, $\chi^2(3) = 15.386, p < .001$.

```
MODEL FIT INFORMATION

Wald Test of Parameter Constraints
```

```
          Value                                  15.356
          Degrees of Freedom                          3
          P-Value                                0.001
```

Intercept Differences

Below is the second omnibus Wald test result for intercept differences. There was evidence that there are significant differences in the distal outcome means across the science attitude classes, $\chi^2(1) = 7.598, p < .001$.

```
MODEL FIT INFORMATION

Wald Test of Parameter Constraints

          Value                                   7.598
          Degrees of Freedom                          1
          P-Value                                0.0058
```

Pairwise Slope and Intercept Differences

To further investigate which class-specific relations differ, pairwise comparisons of the regression slopes and means of the distal outcome are shown below.

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
New/Additional Parameters
    SLOPE12           -0.003      0.062     -0.045      0.964
    SLOPE13           -0.020      0.054     -0.364      0.716
    SLOPE14           -0.096      0.042     -2.259      0.024
    SLOPE23           -0.017      0.074     -0.230      0.818
    SLOPE24           -0.093      0.051     -1.809      0.070
    SLOPE34           -0.076      0.037     -2.076      0.038
    INT12              0.147      0.053      2.756      0.006
    INT14             -0.171      0.045     -3.827      0.000
    INT24             -0.318      0.037     -8.543      0.000
```

Here, `SLOPE12` is the pairwise difference between the slopes in classes 1 and 2. Class 4 (Pro-Science with Elevated Utility) was significantly different from Class 1 (Ambivalent with Elevated Utility) and Class 4 (Ambivalent with Minimal Utility), $p < .05$. Comparisons across intercepts (or the distal outcome means) are all significant.

Slope and Intercept Coefficients

Additionally, each regression between the predictor and outcome can be examined across classes, as well as the intercept coefficients (*Note*: Recall that the mean of the distal outcome factor, Interest in Science Issues, was set to zero for the Ambivalent w/ Minimal Utility Value for measurement identification when adding the latent variable. This class was used as the reference class. Thus, the mean of the factor is set to zero, and others are compared to it.)

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Latent Class 1

 ISSUES   BY
    KA9B               1.000      0.000    999.000    999.000
    KA9D               1.226      0.033     36.595      0.000
    KA9G               1.157      0.034     34.063      0.000
    KA9K               0.783      0.029     27.295      0.000

 ISSUES     ON
    URM                0.050      0.025      1.974      0.048
    FEMALE            -0.156      0.020     -7.892      0.000
    MOTHED             0.007      0.009      0.706      0.480
    KSCIIRT            0.077      0.039      1.989      0.047
!Slope coefficient

 Means
    N#1                0.874      0.000    999.000    999.000
    N#2               -6.329      0.000    999.000    999.000
    N#3               -5.617      0.000    999.000    999.000

 Intercepts
    KA9B               1.842      0.040     45.710      0.000
    KA9D               1.941      0.049     40.006      0.000
    KA9G               1.998      0.046     43.232      0.000
    KA9K               1.733      0.033     52.902      0.000
    ISSUES             0.005      0.052      0.092      0.927
!Intercept coefficient
```

```
 Residual Variances
    KA9B               0.232      0.010     23.948      0.000
    KA9D               0.130      0.008     15.568      0.000
    KA9G               0.149      0.008     17.872      0.000
    KA9K               0.295      0.010     29.088      0.000
    ISSUES             0.189      0.021      9.007      0.000

Latent Class 2

 ISSUES   BY
    KA9B               1.000      0.000    999.000    999.000
    KA9D               1.226      0.033     36.595      0.000
    KA9G               1.157      0.034     34.063      0.000
    KA9K               0.783      0.029     27.295      0.000

 ISSUES     ON
    URM                0.050      0.025      1.974      0.048
    FEMALE            -0.156      0.020     -7.892      0.000
    MOTHED             0.007      0.009      0.706      0.480
    KSCIIRT            0.080      0.051      1.573      0.116
!Slope coefficient

 Means
    N#1               -1.487      0.000    999.000    999.000
    N#2                6.772      0.000    999.000    999.000
    N#3                4.393      0.000    999.000    999.000

 Intercepts
    KA9B               1.842      0.040     45.710      0.000
    KA9D               1.941      0.049     40.006      0.000
    KA9G               1.998      0.046     43.232      0.000
    KA9K               1.733      0.033     52.902      0.000
    ISSUES            -0.143      0.051     -2.801      0.005
!Intercept coefficient

 Residual Variances
    KA9B               0.232      0.010     23.948      0.000
    KA9D               0.130      0.008     15.568      0.000
    KA9G               0.149      0.008     17.872      0.000
    KA9K               0.295      0.010     29.088      0.000
    ISSUES             0.183      0.023      8.049      0.000

Latent Class 3
```

```
 ISSUES   BY
    KA9B               1.000      0.000    999.000    999.000
    KA9D               1.226      0.033     36.595      0.000
    KA9G               1.157      0.034     34.063      0.000
    KA9K               0.783      0.029     27.295      0.000

 ISSUES     ON
    URM                0.050      0.025      1.974      0.048
    FEMALE            -0.156      0.020     -7.892      0.000
    MOTHED             0.007      0.009      0.706      0.480
    KSCIIRT            0.097      0.035      2.781      0.005
!Slope coefficient

 Means
    N#1               -0.606      0.000    999.000    999.000
    N#2                0.392      0.000    999.000    999.000
    N#3                2.431      0.000    999.000    999.000

 Intercepts
    KA9B               1.842      0.040     45.710      0.000
    KA9D               1.941      0.049     40.006      0.000
    KA9G               1.998      0.046     43.232      0.000
    KA9K               1.733      0.033     52.902      0.000
    ISSUES             0.000      0.000    999.000    999.000
!Intercept coefficient (reference class)

 Residual Variances
    KA9B               0.232      0.010     23.948      0.000
    KA9D               0.130      0.008     15.568      0.000
    KA9G               0.149      0.008     17.872      0.000
    KA9K               0.295      0.010     29.088      0.000
    ISSUES             0.174      0.018      9.812      0.000

Latent Class 4

 ISSUES   BY
    KA9B               1.000      0.000    999.000    999.000
    KA9D               1.226      0.033     36.595      0.000
    KA9G               1.157      0.034     34.063      0.000
    KA9K               0.783      0.029     27.295      0.000

 ISSUES     ON
    URM                0.050      0.025      1.974      0.048
```

```
    FEMALE              -0.156       0.020      -7.892       0.000
    MOTHED               0.007       0.009       0.706       0.480
    KSCIIRT              0.173       0.012      14.791       0.000
!Slope coefficient

 Means
    N#1                 -4.425       0.000     999.000     999.000
    N#2                -13.804       0.000     999.000     999.000
    N#3                -10.500       0.000     999.000     999.000

 Intercepts
    KA9B                 1.842       0.040      45.710       0.000
    KA9D                 1.941       0.049      40.006       0.000
    KA9G                 1.998       0.046      43.232       0.000
    KA9K                 1.733       0.033      52.902       0.000
    ISSUES               0.176       0.033       5.282       0.000
!Intercept coefficient

 Residual Variances
    KA9B                 0.232       0.010      23.948       0.000
    KA9D                 0.130       0.008      15.568       0.000
    KA9G                 0.149       0.008      17.872       0.000
    KA9K                 0.295       0.010      29.088       0.000
    ISSUES               0.162       0.010      16.391       0.000
```

End of Annotated Output