

Prediction from latent classes: A demonstration of different approaches to include distal  
outcomes in mixture models

Karen Nylund-Gibson  
University of California, Santa Barbara  
Department of Education  
Santa Barbara, CA 93106  
knylund@education.ucsb.edu

Ryan P. Grimm  
University of Virginia  
Curry School of Education  
Charlottesville, VA 22903  
[rpg5hu@virginia.edu](mailto:rpg5hu@virginia.edu)

Katherine E. Masyn  
School of Public Health, Georgia State University  
Atlanta, GA 30303  
(404) 413-9315  
kmasyn@gsu.edu

Correspondence concerning this article should be address to Karen Nylund-Gibson, Department of Education, University of California, Santa Barbara, Department of Education, Santa Barbara, CA 93106. E-mail: [knylund@education.ucsb.edu](mailto:knylund@education.ucsb.edu).

The research reported here was partially supported in part by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A160157 to the University of California, Santa Barbara to Dr. Nylund-Gibson as co-PI. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

Accepted for publication in *Structural Equation Modeling* on February 24, 2019

## **Abstract**

Including auxiliary variables such as antecedent and consequent variables in mixture models provides valuable insight in understanding the population heterogeneity embodied by a latent class variable. The model building process regarding how to include predictors/correlates and outcomes of the latent class variables into mixture models is an area of active research. As such, new methods of including these variables continue to emerge and best practices for the application of these methods in real data settings (including simple guidelines for choosing amongst them) are still not well established. This paper focuses on one type of auxiliary variable—distal outcomes—providing an overview of the methods currently available for estimating the effects of latent class membership on subsequent distal outcomes. We illustrate the recommended methods in the software packages Mplus and Latent Gold using a latent class model to capture population heterogeneity in students' mathematics attitudes, linking latent class membership to two distal outcomes.

## Introduction

Mixture models are becoming a widely used statistical technique in the social sciences as a person-centered approach to understanding and modeling (unobserved) population heterogeneity. Among the wide range of mixture models are the commonly-used cross-sectional models such as latent class analysis (LCA) and latent profile analysis (LPA), and longitudinal models including growth mixture models (GMM) and latent transition analysis (LTA). All of these models, whether cross-sectional or longitudinal, have one thing in common: they estimate at least one multinomial latent variable that divides the population into a finite number of mutually exclusive and exhaustive latent categories (classes).

While establishing the measurement model for the latent class variable and describing the marginal distribution of individuals across the latent classes are often primary goals of a mixture analysis, researchers are typically interested in understanding more about the consequences of membership in the emergent latent classes by investigating the structural and predictive associations between the latent class variable and subsequent distal outcomes. These distal outcomes are usually conceptualized as consequences of latent class membership rather than indicators or manifestations of latent class membership; that is, the distal outcomes are not considered part of the measurement model within the set of variables believed to reflect and define the latent classes. For example, in a study of the heterogeneity in children's readiness for Kindergarten, Quirk, Nylund-Gibson, and Furlong (2013) linked latent classes of school readiness to later academic outcomes, e.g., standardized reading scores. The distal academic outcomes of this example are substantively and theoretically distinct from the set of indicator variables characterizing the latent classes of readiness.

Although the motivation to study the consequences of latent class membership is quite intuitive from a substantive standpoint, the inclusion of distal outcomes in mixture models from a methodological standpoint is not straightforward. Part of the challenge the various technical approaches have been designed to address is that there is no way analytically to distinguish between latent class indicators and distal outcomes in the likelihood function of a mixture model that simultaneously includes both the measurement model for a latent class variable and a structural predictive relationship connecting latent class membership to a distal outcome. To date, there are six different distal outcome methods, described in detail below, from which a researcher can choose, each with their own advantages and disadvantages given the modeling context.

### **Antecedent and consequent variables in mixture models**

The discussion of how best to include distal outcomes in mixture models is part of a larger and ongoing discussion in the methodological literature about *when* and *how* structural variables, hypothesized to be directly related to latent class membership but ancillary to the latent class variable measurement model, should be included in the modeling process when using mixture models. The inclusion of structural variables in mixture models introduces some complexity because the structural variables may unintentionally change the nature of the heterogeneity being modeled (Nylund-Gibson, Grimm, Quirk, & Furlong, 2014; Vermunt, 2010). That is, the measurement model for the latent class variable (encompassing the characterization of each latent class and who in the population resides in each of those classes) can substantially shift when moving from the unconditional latent class measurement model to a structural equation mixture model including antecedent causes and subsequent effects of latent class

membership if the relationships between the latent class indicators and the antecedents and consequents are misspecified in any way.

A recent simulation study by Nylund-Gibson and Masyn (2016), using latent class analysis, focused on understanding the impact of covariate effect misspecification on the class enumeration process. The results indicate that, regardless of the true covariate relationship with the latent class variable and the latent class indicators, the class enumeration process would consistently result in the correct number of classes when conducted without covariates (e.g., the unconditional latent class measurement model), at a rate comparable to the correctly specified model with covariates. In contrast, a misspecified latent class regression model used for the class enumeration performed poorly. While that study focused on latent class antecedents, they speculated that the recommendation to conduct the latent class enumeration without latent class predictors should extend to the exclusion of distal outcomes during enumeration as well.

For an applied researcher, the shift that may occur in the meaning and composition of the latent classes from the final unconditional latent class measurement model to the conditional structural model can be quite unsettling. It signals a contamination of the measurement model such that variables intended to explain (or be explained by) heterogeneity captured by the latent class variable are instead influencing the character and essence of that latent variable. Many of the newer methods for including distal outcomes attempt to preserve the latent class formation based on the unconditional measurement model while enabling prediction from the latent class variable to the outcome.

### **The Current Paper**

Building from current papers which have provided insight into the methods available for distal outcome estimation (e.g., Clark & Muthen, 2009; Petras & Masyn, 2010), this paper will

provide a detailed overview of the most current methodological landscape related to the incorporation of distal outcomes in mixture models. We have identified six different approaches for including distal outcomes that are available, some of which will be discussed but are not advised for use. All of these methods are available in the software package Mplus (Muthén & Muthén, 1998-2015). We will briefly introduce each method and discuss the logic behind the method and how they are different and then demonstrate the method using real data examples.

As an empirical demonstration, we will use a latent class analysis (LCA; Goodman, 1974; McCutcheon, 1987), a commonly used mixture model that often includes distal outcomes. We consider both a continuous and a categorical distal outcome. Furthermore, we consider the effect of latent class membership on the distal outcome(s) adjusting for observed potential confounders (i.e., directly predictive of latent class membership and the distal outcome). It would be extremely rare in an applied setting not to have any control variables requiring inclusion in the model examining the association between the latent class variable and distal outcome. While we focus on how these distal outcome methods are implemented in Mplus and Latent Gold, some of the methods we will discuss are available in other software (c.f., SAS, Stata, and various R packages) though are not included in this paper.

### **Approaches for Including Distal Outcomes in Mixture models**

With one exception, all of the approaches for including distal outcomes of latent class membership in a mixture model can be characterized as “stepwise” approaches in which either an unconditional latent class measurement model or latent class regression model is first estimated and then information related to latent class membership is extracted for use in a subsequent model to estimate the relationship between latent class membership and the distal outcome. In the following sections, we begin with the exception, i.e., the only non-stepwise (i.e.,

*one step*) approach to including distal outcomes. We then introduce each stepwise approach and provide references for further reading about the statistical model behind each approach. For each, we will provide advantages and disadvantages and, when applicable, we provide the syntax for specifying each in Mplus 8.0 which can be found in the Appendices. We also provide syntax for Latent Gold 5.1 for two of the approaches in the Appendices. We start with stepwise approaches that we classify as “Outmoded Approaches”, meaning these are methods that are not currently supported for use based on findings found in the methodological literature (for reasons we explicate later). We include them, nonetheless, in this paper as a way to contextualize the development and history of the range of distal outcome methods and to provide a rationale for why outdated methods are no longer supported.

### **One-step Approach: Distal-as-indicator**

In the distal-as-indicator approach, the distal outcome is part of the measurement model of the latent variable; that is, the distal outcome is treated as an indicator of the latent class variable (Muthén & Shedden, 1999). This method is sometimes referred to as the one-step approach (Bandein-Roche, Miglioretti, Zeger & Rathouz, 1997) because the relationships between the latent class variable and the observed latent class indicator variables as well as the the class-specific distal outcome distributions are estimated simultaneously in a single analytic model (Vermunt, 2010); thus, *one* modeling step. As a result, the heterogeneity encompassed by the latent class variable is assumed to manifest in the observed joint distribution of the initial set of latent class indicators *and* the distal outcome. In other words, in the likelihood function for the observed data, including the latent class indicators and distal outcome(s), there is nothing to distinguish the indicators from the distal outcomes. As such, the class-specific distribution of the distal outcome cannot be interpreted as representing the direct effect of a latent class variable, as

measured by the other indicators. Because of this, the distal-as-indicator approach to including distal outcomes is unique and not comparable to the other methods discussed below because the latent class measurement model, with the distal-as-indicator, is inherently different than a measurement model that excludes the distal outcome. In fact, if the distal-as-indicator approach is used, then the entire measurement model building process, including the class enumeration, should be done with the distal outcome included with the other latent class indicators (Petras & Masyn, 2010). To be clear, the one-step approach should not be artificially divided into a stepwise approach, e.g., running one-step models with and without the distal outcome(s). One should expect that the number, composition, and meaning of the latent classes will shift when comparing a latent class model including only the indicators to a one-step model including the distal outcomes.

To summarize, the use of this approach must be substantively motivated by a conceptualization of a latent class variable intended to capture underlying heterogeneity in the joint distribution of all the indicators, including the distal outcome(s). The class-specific distributional parameters for the distal outcome (e.g., class-specific means and variances) are in the measurement model parameter set and should be evaluated in terms of measurement properties of the distal-as-indicator, i.e., class homogeneity and class separation with regards to the distal outcome indicator. The resultant latent classes should be labeled and interpreted with consideration of the distal-as-indicator. Direct effects from a covariate to the distal outcome should be interpreted as evidence of measurement noninvariance of the distal-as-indicator. Because this approach uses a different set of measured variables as class indicators (manifest variables and the distal outcome), the latent class variable is qualitatively different in this approach and thus not comparable to the other methods. Because of this, we do not demonstrate



this method in this paper but do provide syntax for the data illustration in Appendix A for reference.

### **Outmoded Stepwise Approaches**

Given the development of auxiliary methods approaches, there are some methods that are no longer supported for use by the methodological literature. These methods, which we label as “outmoded,” have been shown to be biased or unstable in simulation studies. Though their use is no longer supported, in some cases they are being used by researchers, likely because they are not able to keep up with the rapid development of the methods or because the methods are still available in some statistical software packages. Our hope is that, by including them here with a discussion about why they should not be used, we can help to reduce their use by applied researchers.

The remaining approaches all use the same measurement model for the latent class variable, excluding the distal outcome, and the results can be directly compared in the data illustration. Each of these approaches begins with a class enumeration, establishment, and evaluation of the measurement model for the latent class variable.

**Classify-analyze approach.** The classify-analyze approach (Clogg, 1995) is a stepwise modeling process, sometimes called a two-step approach (Step 1: classify; Step 2: analyze) or three-step (Step 1: estimate; Step 2: classify; Step 3: estimate relationships between latent class membership and other variables). In this approach, based on the final estimated unconditional measurement model, individuals are classified into latent subgroups (the “classify” step) and then latent class assignment is used as an observed grouping (i.e., multinomial) variable predictor of the distal outcome in a subsequent model (the “analyze” step). In the “classify” step of this approach, individuals are assigned to a latent class based on their modal posterior class

probability, or modal class assignment (Masyn, 2013). That is, each individual in the sample is fully assigned to the latent class for which they have the highest estimated probability of membership based on their observed data and the estimated Step 1 model. These modal class assignments can be saved into a new data file and merged with other variables. Then, in the “analyze” step, class assignment is used as an observed multinomial grouping variable in an ANOVA, or other analysis appropriate for the scale of the distal outcome. This allows for the investigation of the effect of assigned latent class membership on the distal outcome(s), e.g., testing for differences in means for the distal outcomes across the assigned latent class subgroups. Covariates can be easily controlled for in the “analyze” step.

A major weakness of the classify-analyze approach is that the “analyze” step treats the modal class assignment as an observed, perfectly measured grouping variable, ignoring the non-zero classification error (i.e., measurement error and estimation error in latent class membership) inherent in the modal assignment made during the classify step. Assuming no classification error in the latent class assignment, when the true error is non-zero, can lead to biased point estimates and standard errors for the effect of latent class membership on the distal outcome (Bakk, Tekle, & Vermunt, 2013; Gudicha & Vermunt, 2013). Some argue when entropy is high, such as .80 or higher (Clark & Muthén, 2009; Ram & Grimm, 2009), the classification error could be considered negligible and this approach estimates mean differences without much bias (Bolck, Croon, & Hagnaars, 2004). However, depending on the latent class proportions and degree of class separation, an overall high entropy value can obscure a high rate of classification error in a smaller class.

Bray, Lanza, and Tan (2015) proposed a modification to the classify-analyze approach which they called “inclusive LCA,” to reduce the error associated with modal class assignment.

With this modification, once the unconditional model is chosen, all auxiliary variables are added as covariates and new posterior probabilities are obtained. The “analyze” step is then conducted using the new posterior probabilities. Bray et al. (2015) argued this would reduce classification error as the new posterior probabilities would be conditioned on individuals’ responses to the auxiliary variables as well as the latent class indicators. However, Asparouhov and Muthén (2014a) noted this approach may fail when the variances of the distal outcomes are not equal across latent classes.

The five remaining approaches are all stepwise approaches in that, like the classify-analyze approach, they aim to de-couple the estimation of the latent class measurement model from the estimation of the structural relationships between latent class membership and distal outcomes, but also aim to explicitly account for, in some way, the classification error in latent class membership, rather than ignoring it. Because of the existence of these improved approaches, the classify-analyze approach is now considered contraindicated for the investigation of distal outcomes of latent class membership, regardless of the entropy level. As a result, we do not demonstrate this method in our paper.

**Pseudo-class draw approach.** The pseudo-class draw approach (Asparouhov & Muthén, 2007; Petras & Masyn, 2010; Wang, Brown, Bandeen-Roche, 2005) is another distal outcome method that assigns individuals to latent classes and then analyzes the structural associations between latent class membership and the distal outcomes, but instead of assigning individuals using the modal class assignment, the assignment is done taking advantage of the posterior probability distribution. Latent class membership is essentially multiply imputed for each individual by taking random draws (Mplus uses a default of five) from the multinomial distribution defined by the set of posterior class probability values obtained for each individual

based on the estimated unconditional mixture model and the individual's observed data on the latent class indicators. For each pseudo-class draw for the entire sample, the association between latent class membership and the distal outcome is estimated and then those estimates are combined (similar to a multiple imputation procedure) across the draws. In Mplus, the pseudo-class draw approach is implemented by specifying the command "(e)" after each of the distal outcome variables in the auxiliary syntax (e.g., auxiliary = distal (e)). As the pseudo-class draw procedure has been shown to yield biased and imprecise effect estimates compared to the more recent stepwise approaches for latent class regression models (Petersen, Bandeen-Roche, Budtz-Jorgensen, & Larsen, 2012), it should also be considered contraindicated for use with predictors or distal outcomes of latent classes and we do not demonstrate it in our paper.

The remaining four stepwise approaches are the current alternatives to the distal-as-indicator approach which, when corresponding assumptions are met, can produce consistent estimates of the effects of latent class membership on distal outcomes. Thus, it is only these three approaches that we demonstrate using the real data example.

### **Current Stepwise Approaches**

**ML three-step approach.** This approach was developed using an error-in-variable schema (Vermunt, 2010). We have already documented that the classify-analyze approach can fail when there is non-negligible error in latent class assignment. However, the error in classification can be estimated and incorporated in the distal outcome model, rather than forcing the no-error assumption.

As the name suggests, this method involves three steps. The first step involves identifying the best-fitting unconditional model and saving the posterior probabilities and modal class assignment for that model. In the second step, the estimated conditional probabilities for

modal class assignment given true latent class membership are computed. These quantities, representing the estimated average classification errors for the modal class assignment, are then translated into fixed parameter values describing the direct relationship the latent class variable and the multinomial modal class assignment variable for use in the third step.

For the third step, a new analytic model is specified. The modal class assignment from the first step is used as a single nominal indicator of the latent class variable with class-specific parameters representing the classification error fixed at the values computed in the second step. This process is analogous to fixing the residual variance of a single indicator of a latent factor to correspond to a known or estimated item reliability. Distal outcomes are then included as additional indicators of the latent class variable. One drawback to this approach is that latent class membership may still shift from the first and third step for two reasons: (1) the classification error computed in the second step is the estimated *average* classification error in the sample but the third step assumes that average classification error applied uniformly for all individuals; (2) the distal outcome is specified as an additional indicator of the latent class variable. See Nylund-Gibson, Grimm, Quirk, and Furlong (2014) for a more detailed presentation of this approach.

For the ML three-step, as well as the other current stepwise approaches, a researcher is often faced with a choice in their preferred software about whether to proceed through the steps “manually” or to utilize convenience features in the software that may automate some or all of the steps. In our experience, the manual approach offers far greater flexibility in model specification. The convenience of the automation camouflages the serious restrictions and assumptions imposed within software-generated (i.e., not user-generated) models. For example, the automated ML three-step in Mplus (auxiliary = distal (DU3STEP) or auxiliary = distal

(DE3STEP)) cannot control for covariates when estimating the relationship between the latent class variable and a distal outcome—a specification likely needed in real data settings and easily done with the manual ML three-step. Furthermore, the automated process is quite restricted in terms of the types of distal outcomes that may be considered, with regards to measurement scales, distributional assumptions, latent outcomes, etc. Similarly, Latent Gold can accommodate either covariates or distal outcomes using the graphical user interface (GUI; Step3 menu), but syntax is required when analyzing both simultaneously. Given these limitations, we recommend that the automated stepwise procedures be reserved for exploratory or simulation purposes only.

The details of the execution of a manual ML three-step approach are necessarily software-specific. For example, in Mplus, including all distal outcomes and covariates in the “auxiliary” variable option and using the “savedata: save = cprob” feature when running the Step 1 model will generate a new data file that can be used for estimating the Step 3 model. Additionally, in Mplus v7.1 and later, Mplus computes the values corresponding to the classification error of the modal class assignment as part of the default output in the Step 1 model to be used as fixed class-specific multinomial intercepts for the modal class assignment indicator in Step 3, found under the section titled “Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Row) by Latent Class (Column),” saving the user from having to do much of the work of Step 2. We provide Mplus syntax for the manual ML three-step using the data illustration in Appendix B (for Step 1) and Appendix C (for Step 3) and Latent Gold syntax in Appendix F.

**BCH Method.** First introduced by Bolck, Croon, and Hagenaars (2004) for categorical covariates, this method was referred to as the BCH method after the last names of the authors of the paper. This method was later modified to accommodate continuous covariates (Vermunt, 2010) and eventually extended to include covariates and distal outcomes (Bakk, Tekle, & Vermunt, 2013) and is also conducted in three steps. The BCH method is very similar to the ML three-step approach except that instead of calculating the average classification error in the second step, classification errors for each individual are computed, and the inverse logits of those individual-level error rates are used as weights in the third step rather than using the modal class assignment as imperfect latent class indicator. Advantages with this approach are that it appears to be more resistant to shifts in the latent class membership from Step 1 to Step 3, compared to the ML three-step and can often be used irrespective of distal outcomes variances being equal or unequal across latent classes (which can be a problem with the LTB approach – see below). However, if entropy is low or if sample size is small, the weights may take on negative values, which can be problematic. In such cases, it is possible for the variances of the distal outcome to take on negative values when variances are class-specific. However, if variances are constrained to equality, this problem will not occur. Thus, if class-specific variances are not of interest, the latter approach is preferred. This situation is more problematic when a more complex model is specified to differ across classes compared to situations in which only distal outcomes are analyzed.

As with the ML three-step approach, researchers will have a choice between the automated (auxiliary = distal (BCH)) or manual BCH three-step process in Mplus and using the GUI (Step3 menu) or syntax specification in Latent Gold. The automated version of the BCH three-step approach are similarly restrictive (see Asparouhov and Muthén, 2014a, for details. The

way in which the BCH method is currently implemented in Mplus precludes us from estimating models that have correlated residuals of categorical distal outcomes, however it is possible to correlate the residuals of continuous distal outcomes. Additionally, the current implementation will only work with a single latent class variable.

As before, the details of the execution of a manual BCH three-step approach are necessarily software-specific. For example, in Mplus, including all distal outcomes and covariates in the “auxiliary” variable option and using the “savedata: save = bchweights” feature when running the Step 1 model will generate a new data file that can be used for estimating the Step 3 model. The “training” option with “(BCH)” is used in the Step 3 model to identify the BCH weight variables carried over from the Step 1 saved data file. Mplus syntax for the manual BCH three-step using the data illustration in Appendix D (for Step 1) and Appendix E (for Step 3) and Latent Gold syntax in Appendix G.

**LTB Approach.** The approach presented in Lanza, Tan, and Bray (2013), and referred to as the LTB approach in Bakk and Vermunt (2016), is also conducted in a stepwise fashion. However, after identifying the optimal unconditional model, it estimates the relationship of the distal outcomes and the latent class variable by performing a latent class regression, thus treating the distal outcome as a covariate. By applying Bayes’ Theorem to the latent class regression, a kernel density estimate for the marginal distribution of the distal outcome, and the estimated marginal distribution of the latent class variable, this approach computes the conditional distribution of the distal outcome given latent class membership, usually in the form of class-specific means of the distal outcomes (Lanza, Tan, & Bray, 2013). This approach is automated in Mplus through the “auxiliary = distal (DCON)” or “auxiliary = distal (DCAT)” options. The manual version of the LTB approach is not nearly as straightforward as for the ML and BCH



three-step approaches as it requires estimation of the marginal distribution of the distal outcome. This approach relies on a key assumption, specifically it assumes the variances of continuous distal outcomes are equal across latent classes. If this is not the case, mean estimates of the distal outcomes may be biased (Bakk & Vermunt, 2016). Further, this approach may not work well if there is low separation between the latent classes (Bakk & Vermunt, 2016). As implemented, it also assumes that the distal outcome(s) are independent of any latent class predictors, conditional on latent class membership. Because of these limitations in the currently available and user-accessible implementations, this approach is not demonstrated in the current paper.

**Two-step method.** The most recently developed stepwise approach is the two-step method (Bakk & Kuha, 2017). Similar to the three-step approaches, the two-step method attempts to correct biased estimates associated with one-step approaches and outmoded stepwise approaches. However, unlike the three-step approaches described above, the two-step method does not attempt to correct the misclassification bias. Rather, once the preferred measurement model is identified, the two-step method fixes the measurement parameters for the latent class variable, while estimating the remaining structural parameters of the model. It is worth noting that while including both covariates and distal outcomes has not been discussed in the literature or included in software packages yet, this two-step approach is capable of accommodating both.

When a distal outcome is the target of analysis, then the measurement parameters from the latent class indicators as well as the distribution of the latent class variable can be estimated in the first step (if there are not covariates). These are all treated as fixed when subsequently estimating the distal outcome in the second step. When analyzing covariate effects, the distribution of the latent class variable from the first step is discarded and re-estimated in the

second step and only the measurement parameters from the latent class indicators are treated as fixed.

A major advantage of this approach is that since measurement parameters are treated as fixed, multiple latent class variables can be included in a single model. Bakk and Kuha (2017) recommend estimating each latent class variable separately. All of these measurement parameters would then be fixed in Step 2, while estimating the structural parameters. Another advantage is that each step can be estimated using different samples. This is especially helpful in terms of missing data. It is not uncommon for distal outcomes to have a lower response rate compared to the data used to estimate a measurement model based on data gathered at an earlier timepoint. Conversely, it is possible that additional distal outcome data may be collected after collecting data on the latent class indicators. The two-step method allows all data from both scenarios to be included. It is important to note there are important statistical details omitted here with respect to implementation, the primary of which is the necessary adjustment of the standard errors in the Step 2 model to account for parameters being fixed to estimated values from the Step 1 model. This adjustment is not trivial and is not currently automated in any software package that we know of. Interested readers should consult Bakk and Kuha (2017) for a more technical explication of this approach.

### **Recommended Approaches to the Inclusion of Distal Outcomes**

In summary, despite the overwhelming and continuously growing list of available approaches to the inclusion of distal outcomes in mixture models, the list of recommended current approaches for real data analysis, based on practical and statistical considerations is relative short (Table 1). The classify-analyze and pseudo-class draw approaches are not recommended under any circumstances. The LTB approach as well as the automated ML three-

step and automated BCH three-step are not recommended as the restrictions and assumptions are unlikely to correspond to any defensible population model. The newest two-step cannot be currently recommended because it is not yet readily accessible to applied researchers. That leaves only two recommended stepwise approaches: (1) the manual ML three-step and (2) the manual BCH three-step. The statistical and practical distinctions between the two have been previously elaborated. It is important to note that neither of these two recommended approaches guarantee that the latent classes from the Step 1 model will be comparable to the latent classes in the Step 3 model; in other words, the inclusion of the distal outcome(s) in Step 3 could still influence the class formation. Using the manual approach, the onus is on the researcher to compare the classification of individuals across the Step 1 and Step 3 models before interpreting the Step 3 latent classes as if they were the same as the Step 1 latent classes and interpreting the difference in the class-specific distributions of the distal outcome(s) in Step 3 as “effects” of latent class membership (as defined in Step 1). Additionally, one must attend to differential patterns of missing data on the latent class indicators and distal outcomes as the analytic sample can unintentionally change from the Step 1 to Step 3 models, depending on the software’s handling for the missingness.

The one-step (distal-as-indicator) approach is conceptually different, as it conceives of the indicators *and* distal outcome(s) as joint manifestations of an underlying latent class variable—the use of this approach should be based on substantive theory and research aims.

### **Demonstration of Recommended Approaches**

As a way of understanding how the recommended methods may (or may not) differ in a real data setting, we demonstrate the use of selected current stepwise approaches to including distal outcomes; specifically, the ML and BCH three-step approaches. We also demonstrate the

one-step approach. However, because the heterogeneity modeled by the distal-as-indicator approach is qualitatively different from the heterogeneity without the distal outcomes, we consider the results of this method separate from the others, though some general comparisons are made. The example we use in this paper is based on the analysis results presented in Masyn (2017). That paper focused on differential item functioning (DIF) of the LCA solution with respect to gender, thus the current paper builds off those results and adds distal outcomes. While simply demonstrating these approaches with a single empirical example does not prove which method is “best,” it is instructive to see the difference in the processes and software specification. One’s own particular research questions and data conditions must inform the selection of one approach over another in a specific analytic context.

Appendices A-E provide the Mplus syntax for: (A) one-step; (B) manual ML three-step Step 1; (C) manual ML three-step Step 3; (D) manual BCH three-step Step 1; and (E) manual BCH three-step Step 3. Though Latent Gold 5.1 (Vermunt & Magidson, 2016) does have a GUI, the analyses conducted in this paper required the use of syntax. Appendix F provides relevant syntax for the ML three-step approach and Appendix G provides the relevant syntax for the BCH three-step approach. Appendices F and G include syntax only for the “variables” and “equations” sections of a syntax file as these are the only sections that must be written manually, while the “options” section can be completed using the GUI. This paper used modal class assignment for the ML and BCH approaches, but proportional assignment can be utilized in Latent Gold by either clicking this option in the GUI, or replacing “modal” with “proportional” in the syntax.

## **Method**

### **Data**

For the data demonstration in this paper we are using a subset of mathematics attitudinal variables from the Longitudinal Survey of American Youth (LSAY; Miller, 1987-1994, 2007). The LSAY sample was drawn from schools across the United States and was stratified by geographic region and urbanicity and consisted of two cohorts. Initial data collection began in 1987, though we use Cohort 2 which began in the fall of 1998. The sample in Cohort 2 was predominantly white (69.5%), followed by African-American (11.2%), Latino/a (9.1%), Asian (3.6%), Native American (1.5%) and 5.1% missing. The numbers of females and males were nearly equal with 48% and 52%, respectively.

**Measures.** Five math attitudinal variables were used as the response indicators for the LCA model. These items were selected from the Grade 8 wave of data collection. These items measured students' attitudes toward math, such as "*I enjoy math*" and "*Doing math often makes me nervous or upset.*" The five self-report items were originally measured on a five-point, Likert-type scale (1 =strongly agree; 2 = agree; 3 = not sure; 4 = disagree; 5 =strongly disagree), but for this analysis we created dichotomous items after reverse coding certain items so that all item endorsements (indicated by value of 1) represented pro-mathematics responses. Table 2 presents the item wording for all five math attitudes, frequencies, and the overall rate of endorsement.

**Covariates and Distal Outcomes.** We used students' self-reported gender as the single covariate in this example. The gender variable, labeled *female*, was coded female =1 for individuals who reported that they were female and female = 0 for those who reported they were male (gender was measured as binary by LSAY). In addition, two distal outcomes were used, one binary and one continuous. Students' feelings about the utility of math in a later job, which we labeled *math-job*, and was coded *math-job* = 1 if the student reported that they felt math was

useful for a later job and 0 = if they did not. Students' math IRT score in grade 9 was used as a continuous distal outcome where higher math scores represent higher math achievement.

### **Analysis Plan**

Class enumeration for these five items is detailed in the Masyn (2017) paper. Once the unconditional model was identified, DIF was explored with respect to gender. For the distal comparisons discussed in this paper, we begin with a 4-class LCA with DIF based on the Masyn (2017) paper. We included the distal outcomes into this LCA model using the distal outcome models described above. All models were fit in Mplus 8.0 (Muthén & Muthén, 1998-2018).

**Enumeration for Distal-as-Indicator.** Since we based our unconditional model on the Masyn (2017) paper, the only class enumeration conducted was for the one-step distal outcome method of distal-as-indicator. For this model, the five math attitude variables, the gender covariate, and the two distal outcomes are included as class indicators. For enumeration, we followed the recommendations of Masyn (2013) and fit sets of models with different numbers of classes. Fit statistics for all models were recorded and compared but were examined in tandem with substantive considerations. We considered information criteria including the Akaike Information Criterion (AIC; Akaike, 1987), Bayesian Information Criterion (BIC; Schwartz, 1978), adjusted BIC (ABIC; Sclove, 1987), Consistent Akaike's Information Criterion (CAIC; Bozdogan, 1987), and Approximate Weight of Evidence Criterion (AWE; Banfield & Raftery, 1993). We also considered two fit statistics derived from information criteria. The Bayes Factor (BF) provides a pairwise comparison between two competing models and the correct model probability (cmP) compares each of  $J$  models to a set of  $J$  models. Finally, we considered two likelihood ratio based tests. The adjusted Lo-Mendell-Rubin likelihood ratio test (LMR; Lo, Mendell, & Rubin, 2001) and the bootstrapped likelihood ratio test (BLRT; McLachlan & Peel,

2000), which provide a  $p$ -value comparing a  $k - 1$  class model to a  $k$  class model with a significant  $p$ -value providing evidence for the  $k - 1$  class model.

**Stepwise Inclusion of Distal Outcomes.** Subsequent to examining DIF, we implemented the recommended stepwise distal outcome procedures outlined above; specifically, the manual ML three-step and the manual BCH three-step. The Step 1 models for each approach accounted for the covariate as a predictor of latent class membership and a source of DIF.

## Results

The results section is divided into subsections. We first present the descriptive statistics for all variables used to enumerate the classes in the LCA model and the distal outcome variables. Next, we present the final 4-class model based on the Masyn (2017) results, followed by the class enumeration results for the distal-as-indicator model because the enumeration process for that model was different than the standard LCA that utilized stepwise approaches for including distal outcomes (discussed more in the relevant section). Finally, we compare the results of the distal outcome mean comparisons using the distal-as-indicator approach, which is different than the stepwise approaches.

### Descriptive Statistics

Table 2 presents the sample size, frequency, and relative frequency for the five math attitude variables, the covariate of gender, and the two distal outcomes that were both measured in 9<sup>th</sup> grade (math IRT achievement scores, importance of math for later job).

### Latent Class Analysis (LCA) of Math Attitudes

The example we use here is based on the data example from Masyn (2017), which explored heterogeneity in students' math attitudes in eighth grade by fitting a series of LCA models with varying number of classes. For the sake of brevity, we do not present the entire class

enumeration process but readers are encouraged to consult Masyn (2017) for details. The LCA on these five binary items of math attitudes resulted in a four-class unconditional model. The item probability plot is presented in Figure 1, which depicts four latent classes labeled class 1 - 4 and were differentiated by their probability of endorsing the different math attitudinal items. After settling on the best unconditional model, DIF was explored.

**Differential Item Functioning (DIF).** After choosing the unconditional model for these five items, Masyn (2017) explored if there was evidence of DIF with respect to gender. There was evidence of DIF, specific to the indicators “enjoyment” and “understanding.” However, Masyn specified two types of DIF, uniform and non-uniform. Uniform DIF refers to a covariate effect on a given indicator that is consistent across latent classes. Non-uniform DIF refers to a covariate effect on a given indicator that varies across latent classes. Masyn found evidence of a DIF effect of gender on “enjoyment” to be uniform and the DIF effect of gender on “understanding” to be non-uniform. When exploring distal outcome differences, the DIF that emerged with respect to gender was incorporated into the modeling of the distal outcomes. Specifically, the Step 1 model used to create the modal class assignment, estimate the classification error rates for the ML three-step, and estimate the BCH weights for the BCH three-step, included gender as a predictor of latent class membership and a source of DIF on “enjoyment” and “understanding”.

### **Comparison of Distal Outcome Approaches**

After choosing the best-fitting unconditional LCA model and specifying the correct DIF with respect to the gender covariate, we proceeded to fit three conditional models, each using one of the aforementioned approaches for including distal outcomes. Both distal outcomes were included in each conditional model.



**Distal-as-indicator approach.** We present the distal estimation results for the distal-as-indicator approach separate from the other results since the presence of the distal outcomes as a source of information slightly changed the nature of the heterogeneity being modeled. Overall, the results from the distal-as-indicator approach suggested a 4-class model as well (see Table 3 for fit information) even when the distal outcomes and gender covariate were included in the model. Though there were some differences in the emergent latent classes compared to the classes from the unconditional LCA model without distals, the substantive interpretation would largely remain the same with respect to the math attitudinal variables.

The class-specific means and significance tests of all pairwise comparisons for math-job and 9<sup>th</sup> grade math IRT score are presented in the distal-as-indicator rows of Table 4. We briefly describe this table because, to our knowledge, this is a novel method of presenting distal outcome results. For both the top and bottom panels of the table, each distal outcome is presented in the leftmost column, the distal outcome approaches are presented in the second column, the latent classes and their class-specific means are presented in the third column, and the pairwise comparisons are presented in the remainder of the body of the table. Differences for each pairwise comparison are presented in each table's cells as the mean for a latent class in the top row minus the mean for a latent class in the third column.

In the distal-as-indicator approach, class 4 had the greatest probability of endorsing the math-job variable, and this was significantly greater than the class-specific means of the other three latent classes. The mean for class 3 was also significantly greater than the means of classes 1 and 2. There was not a significant difference between the class 1 and class 2 means. The findings were similar concerning 9<sup>th</sup> grade math IRT score, with two exceptions. The mean math

IRT scores for classes 1 and 3 were not significantly different, and the mean math IRT scores for classes 1 and 2 were significantly different.

We do not make direct comparisons between the distal-as-indicator approach and the remaining two approaches because the distal-as-indicator approach includes the covariate and distal outcomes in the measurement model, essentially rendering it incomparable to the other models, which only include the math attitude indicators in the measurement model.

**Comparison of manual ML and BCH three-step methods.** Class-specific means for the binary math-job variable were largely similar across the two approaches. For instance, the BCH approach yielded a mean (proportion) of .57 for class 1, and the manual 3-step yielded a mean of .58. The largest mean discrepancy across the two approaches occurred with class 2 (.54 in BCH and .50 in the manual 3-step). Most of the results concerning pairwise mean comparisons were similar across both approaches. For instance, class 4 had the greatest probability of endorsing math-job and this was significantly greater than the other three classes; the mean of class 3 was significantly greater than the mean of class 2 and there was no significant difference between the means of classes 1 and 2. However, there was one notable incongruity. Specifically, in the BCH approach, the means for class 3 and class 1 were significantly different, but this was not the case in the manual ML three-step.

The results for the 9<sup>th</sup> grade math IRT distal outcome were very similar for the manual BCH and ML three-step approaches. When comparing corresponding classes across approaches, there was little difference in the estimated means. For example, the greatest difference occurred with class 3 (57.49 in BCH and 57.32 in manual 3-step) and the remaining differences among corresponding classes across approaches were .11 or less. The patterns of statistically significant pairwise comparisons were also similar across approaches. Class 4 achieved a significantly

greater 9<sup>th</sup> grade math IRT score compared to the other three classes; class 3 scored significantly higher than class 1, but not class 2; the scores for classes 1 and 2 were not significantly different.

It should be noted that since we did estimate both covariate and distal outcomes, including the regression of the distal outcome on the covariates (as depicted in Figure 2), the latent class differences in the distal outcome means are adjusted for gender.

## **Discussion**

The use of distal outcomes in mixture models provides a context to understand consequences of membership in the emergent latent classes. While the interest in knowing if there are differences in distal outcome means across classes is quite intuitive, the specification of these mean differences is not that straightforward and is an area of active research. Among the different methods currently available for estimating distal outcome effects, each has advantages and disadvantages. We summarize the current methodological status of each of these methods, provide recommendations based on statistical and practical considerations, and demonstrate the use of the recommended approaches. While the results of our application of these methods with a given dataset do not constitute a methodological evaluation of the performance of these approaches under various population conditions (as would a simulation study), our hope is that including an example provides an opportunity for applied researchers to see software syntax for each approach as well as an example of how to interpret the results. While we have summarized what we believe to be the state-of-the-art for including distal outcomes in mixture models in the moment, this remains an area of active research and future readers should anticipate new methodological evaluations, developments, and changing best practices following the publication of this paper.

### **Current Recommendations Based the Methodological Literature**

The current methodological literature recommends against the use of the classify-analyze and the pseudo-class draws approaches, countermanding any previous recommendations to the contrary. The automated ML and BCH three-step approaches as well as the LTB approach, as currently implemented, are prohibitively restrictive with regards to the nature of the distal outcomes which can be considered and the assumptions regarding the relationships between the predictors and outcomes of latent class membership. As implemented currently in software, the manual ML three-step approach is the most flexible in that it allows for covariate effects on the latent class variable and distal outcome(s) to be estimated simultaneously as well as the inclusion of multiple latent class variables (e.g., latent transition analysis) and it accounts for classification error associated with non-perfect assignment of individuals to latent classes. While we do acknowledge that the manual ML three-step is the most complex to implement for the applied research, involving multiple input files and detailed copying-pasting of fixed parameter values, we hope the challenges in execution are not a barrier to use. We advise researchers to simply proceed with care to avoid syntactical mistakes that may produce erroneous results.

In our experience, the manual BCH three-step approach also performs well with continuous distal outcomes. In Mplus, the manual BCH version also allows for the simultaneous inclusion of covariates and distal outcomes, but is limited to one latent class variable, and a distal outcome must be present (i.e., it cannot be used with a covariate alone). In addition, with the manual BCH it is possible to specify a secondary model, such as class-specific regressions. The ML and BCH three-step approaches are robust when class-specific variances of the distal outcome are not equal, but specifying equal variances is preferred when using the BCH method. In Latent Gold, using the syntax, both the ML and BCH three-step approaches are able to estimate complex models that include residual correlations among the indicators.

While the BCH seems to hold a lot of promise, there are a few limiting factors that emerge in real data contexts. Specifically, the BCH weights may become unstable—that is, it negative BCH weights—thus class-specific distal outcome means are not able to be computed. This issue may be related to the measurement model of the latent class variable and may be more common in more complex latent class measurement models such as growth mixture models (GMM). Further, the BCH method is limited to modeling contexts that include only one latent class variable, which is true in both Mplus and Latent Gold, thus it is not possible to use BCH weights with latent transition analysis models (though you can use the manual MLthree-step in those contexts; see Nylund-Gibson, et al., 2014 for an example).

The inclusion of auxiliary variables, both covariates and distal outcomes, is currently an active area of methodological research. There is still much to be learned, including more about the performance of, and possible limitations of, the methods discussed in this paper. In addition, with the more recent focus on these methods, there likely will be further development of new methods for including auxiliary variables into mixture models. As our understanding advances, we may establish more precise recommendations for use with particular mixture models applied to particular types of data. For example, in a GMM it may be that the growth factors themselves, instead of, or in addition to, the latent class variable, are directly related to the distal outcome. In this situation, the current three-step methods would not be appropriate and perhaps respecification of the GMM to a latent class growth model or use of the two-step method may offer an advantage.

### **Future directions**

The field is at an exciting time of development, with new and innovative methods rapidly being developed, which allow us to include auxiliary variables into mixture models that preserve

the measurement qualities of the latent class variable while also estimating relations amongst study variables. However, these methods are being developed faster than they can be implemented into commonly used software packages. While rapid development inspires rapid adoption, there are many gaps in our understanding regarding the boundaries and limitations of the use of these approaches in practice. It is often the case that any limitations we may encounter as researchers are not of the method itself but of the way the method is implemented in a software package. Additionally, as methods for the inclusion of auxiliary variables into mixture models continue to be developed, we need to ensure they work for the range of mixture models and a variety of auxiliary variables. For example, the newly proposed two-step method looks very promising in terms of its flexibility and generalizability to multiple types of mixture models. Once this approach is integrated into more software packages, it will be imperative to test its performance under a variety of conditions and model specifications. We hope our paper provides a clear portrait of this one methodological moment in time, guiding researchers actively applying mixture model in the near-future and, in the more distant future, becomes a part of the historical backdrop for the new approaches as they emerge.

## References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317–332. doi: 10.1007/BF02294359
- Asparouhov, T. & Muthén, B.O. (2007). Wald test of mean equality for potential latent class predictors in mixture modeling. From <http://www.statmodel.com/download/MeanTest1.pdf>
- Asparouhov, T., & Muthén, B. (2014a). *Auxiliary variables in mixture modeling: Using the BCH method in Mplus to estimate a distal outcome model and an arbitrary secondary model* (Mplus Web Note: No. 21). Retrieved from: [http://www.statmodel.com/download/asparouhov\\_muthen\\_2014.pdf](http://www.statmodel.com/download/asparouhov_muthen_2014.pdf)
- Asparouhov, T., & Muthén, B. (2014b). Auxiliary Variables in Mixture Modeling : 3-Step Approaches Using Mplus. *Structural Equation Modeling: An Interdisciplinary Journal*, 21, 329-341. doi: 10.1080/10705511.2014.915181
- Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, 92(440), 1375-1386.
- Bakk, Z., & Kuha, J. (2017). Two-step estimation of models between latent classes and external variables. *Psychometrika*, doi: 10.1007/s11336-017-9592-7
- Bakk, Z., Tekle, F.B., & Vermunt, J.K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43, 272-311. doi: 10.1177/0081175012470644
- Bakk, Z. Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: An Interdisciplinary Journal*, 23(1), 20-

31. doi: 10.1080/10705511.2014.955104
- Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*, 3-27. doi: 10.1093/pan/mp001
- Bray, B.C., Lanza, S.T., & Tan, X. (2015). Eliminating bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling: An Interdisciplinary Journal, 22*, 1-11. doi: 10.1080/10705511.2014.935265
- Clark, S., & Muthén, B. (2009). *Relating Latent Class Analysis results to variables not included in the analysis*. Retrieved from <http://www.statmodel.com/download/relatinglca.pdf>
- Clogg, C.C. (1995). Latent class models: Recent developments and prospects for the future. In: Arminger G, Clogg CC, Sobel ME, editors. *Handbook of statistical modeling for the social and behavioral sciences*. New York, NY: Plenum Press. pp. 311–359.
- Gudicha, D. W., & Vermunt, J. K. (2013). Mixture model clustering with covariates using adjusted three-step approaches. In *Algorithms from and for Nature and Life* (pp. 87-94). Springer.
- Lanza, S.T., Tan, X., & Bray, B.C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling, 20*, 1-26. doi: 10.1080/10705511.2013.742377
- Masyn, K.E. (2013). Latent class analysis and finite mixture modeling. In T. Little (Ed.), *Oxford Handbook of Quantitative Methods* (pp. 551-611). New York: Oxford University Press. Doi: 10.1093/oxfordhb/9780199934898.013.0025



- Masyn, K.E. (2017). Measurement invariance and differential item functioning in latent class analysis with stepwise multiple indicator multiple cause modeling. *Structural Equation Modeling, 24*, 180-197. doi: 10.1080/10705511.2016.1254049
- Miller, J.D. (1987-1994, 2007). *Longitudinal Study of American Youth* [Data file and code book]. Ann Arbor, MI: Inter-university Consortium for Political and Social Science Research.
- Miller, J., and Kimmel, L. (2007). A Longitudinal Model of Educational Attainment and Career Development. A paper presented to the annual meeting of the American Educational Research Association, Chicago, Illinois, April 9, 2007.
- Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*, 369-377.  
doi: 10.1037/1082-989X.8.3.369
- Muthén, L.K., & Muthén, B.O. (1998-2015). *Mplus user's guide*. Los Angeles: Muthén & Muthén,
- Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*(2), 463-469.
- 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. (S. Otten, K. Sassenberg, & T. Kessler, Eds.)*Structural Equation Modeling, 14*(4), 535–569. doi:10.1080/10705510701575396
- Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A Latent Transition Mixture Modeling Using the Three-Step Specification. *Structural Equation Modeling, 21*, 439-454. doi: 10.1080/10705511.2014.915375

- Nylund-Gibson, K., & Masyn, K. (in press). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effect on class enumeration. *Structural Equation Modeling*. doi:10.1080/10705511.2016.1221313
- Petersen, J., Bandeen-Roche, K., Budtz-Jørgensen, E., & Larsen, K. G. (2012). Predicting latent class scores for subsequent analysis. *Psychometrika*, 77(2), 244-262. Doi: 10.1027/1614-2241/a000114
- Petras, H. & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A.R. Piquero, & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 69–100). New York, NY: Springer. doi: 10.1007/978-0-387-77650-7\_5
- Quirk, M., Nylund-Gibson, K., Furlong, M. (2013). Exploring patterns of Latino/a children's school readiness at kindergarten entry and their relationship with Grade 2 achievement. *Early Childhood Research Quarterly*, 28, 437-449. doi: [10.1016/j.ecresq.2012.11.002](https://doi.org/10.1016/j.ecresq.2012.11.002)
- Ram, N. & Grimm, K.J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, 33, 565-576. doi: 10.1177/0165025409343765
- Sclove, L. (1987). *Application of model-selection criteria to some problems in multivariate analysis*. *Psychometrika*, 52(3), 333–343. doi: 10.1007/BF02294360
- Wang C.P., Brown, C.H., Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association*, 100(3), 1054-1076. doi: [10.1198/016214505000000501](https://doi.org/10.1198/016214505000000501)

Vermunt, J. K. (2010). Latent Class Modeling with Covariates: Two improved Three-Step Approaches. *Political Analysis*, 18, 450–469. doi:10.1093/pan/mpq025

Vermunt, J.K., & Magidson, J. (2016). Upgrade manual for Latent GOLD 5.1. Belmont MA: Statistical Innovations Inc.

Table 1.

*Summary of Existing Distal Outcome Approaches with Recommendations for Use*

Recommendation	Approach	Comments
Recommended for informed selection and use.	Manual ML three-step	See text for elaborated statistical comparison as well as comparison of available implementations in existing software.
	Manual BCH three-step	
	One-step (distal-as-indicator)	Conceptually-distinct from stepwise approaches; Treats distal outcome as additional indicator of latent class variable.
Recommended, though with some reservations.	Automated ML three-step	Severe restrictions and untenable/untestable assumptions related to distal outcome distribution and relationships between distal outcomes and predictors; Will work well, statistically-speaking, if your data-generation process matches the restrictions and assumptions but that match is, practically-speaking, unlikely.
	Automated BCH three-step	
	LTB	
	Two-step	Promising but (currently) no straightforward way to implement in existing software.
Not recommended under any circumstances.	Classify-analyze	Statistically contraindicated; Historical predecessors to other approaches; Were formally part of mixture modeling best-practices but are now ill-advised.
	Pseudo-class draws	

Table 2.

*Descriptive Statistics of Math IRT Scores, Distal Outcomes, and Covariate*

Variable	Item label	<i>n</i>	<i>f</i>	<i>rf</i>
<b>Math Attitude Indicator</b>				
I enjoy math.	enjoym <sup>a</sup>	2,668	1784	0.67
I am good at math.	goodm <sup>a</sup>	2,670	1850	0.69
I usually understand what we are doing in math.	understandm <sup>a</sup>	2,648	2020	0.76
Doing math often makes me nervous or upset.	nervousm <sup>b</sup>	2,622	1546	0.59
I often get scared when I open my math book and see a page of problems.	scaredm <sup>b</sup>	2,651	1821	0.69
<b>Covariates and Distal outcomes</b>				
Gender (binary: 1-female; 0-male)	female	3,116	1496	0.48
Math is important for a good job (binary: 1-Agree; 0-Disagree)	math-job <sup>a</sup>	2,321	1601	0.69
9th grade Math IRT score	math IRT	2,241	58.81 <sup>c</sup>	12.60 <sup>d</sup>

*Note.* <sup>a</sup>Dichotomized: 1 = strongly agree, agree; 0 = not sure, disagree, strongly disagree. <sup>b</sup>Dichotomized: 1 = strongly disagree, disagree; 0 = not sure, agree, strongly agree. <sup>c</sup>9th grade Math IRT mean. <sup>d</sup>9th grade Math IRT standard deviation. *N*= sample; *f* = frequency; *rf* = relative frequency

Table 3.

*Fit Statistics for the Distal-as-Indicator LCA model (5 Math Attitudes and 2 Distal Outcomes as Class Indicators)*

# classes	# param	LL	AIC	BIC	ABIC	LMR <i>p</i> -value	BLRT <i>p</i> -value	CAIC	AWE	BF	cmP
1	9	-20669	41356.7	41411.1	41382.5	—	—	38738.6	38788.1	0.0	0
2	18	-19656	39347.2	39456.0	39398.9	—	—	39392.1	39491.0	< 1	<.001
3	27	-19421	38895.5	39058.7	38972.9	<.001	<.001	38962.8	39111.1	<1	<.001
4	36	-19349	38770.2	38987.8	38873.4	<.001	<.001	38860.0	39057.7	149828220.6	1
5	45	-19332	38753.5	39025.5	38882.5	0.299	<.001	38865.7	39112.9	60008693.4	<.001
6	54	-19313	38734.9	39061.3	38889.7	—	—	38869.5	39166.2	—	<.001

*Note.* LL = Log-likelihood; npar = number of parameters; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; aBIC = Adjusted Bayesian Information Criterion; CAIC = Consistent Akaike Information Criterion; AWE = Approximate Weight of Evidence Criterion; BLRT = Bootstrapped Likelihood Ratio Test; LMR = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test; BF = Bayes Factor. Cells with “—” indicates that the value was not estimated.

Table 4.

*Significant Differences Between Class-Specific Means of Both Distal Outcomes of Math-Job and Math Achievement by the Different Distal Outcome Methods*

	Distal Approach	Class (distal mean)	c1	c2	c3
Math-job	One-step (distal-as-indicator)	c1 (.57)	0		
		c2 (.56)	0.01		
		c3 (.64)	-0.07*	-.008*	
		c4 (.79)	-0.22***	-0.23***	-.015***
	Manual BCH three-step	c1 (.57)	0		
		c2 (.54)	0.03		
		c3 (.69)	-0.12**	-0.15**	
		c4 (.77)	-0.20***	-0.23***	-0.08*
	Manual ML three-step	c1 (.58)	0		
		c2 (.50)	0.08		
		c3 (.66)	-0.08	-0.16**	
		c4 (.79)	-0.21***	-0.29***	-0.13*
Math Achievement	One-step (distal-as-indicator)	c1 (53.00)	0		
		c2 (56.29)	-3.29*		
		c3 (52.48)	0.52	3.81*	
		c4 (63.44)	-10.44***	-7.15***	-10.96***
	Manual BCH three-step	c1 (52.77)	0		
		c2 (55.35)	-2.59		
		c3 (57.49)	-4.72***	-2.14	
		c4 (62.71)	-9.94***	-7.36***	-5.22***
	Manual ML three-step	c1 (52.72)	0		
		c2 (55.46)	-2.74		
		c3 (57.32)	-4.59***	-1.85	
		c4 (62.76)	-10.04***	-7.30***	-5.45***

\*  $p < .05$ , \*\*  $p < .01$  \*\*\*  $p < .001$

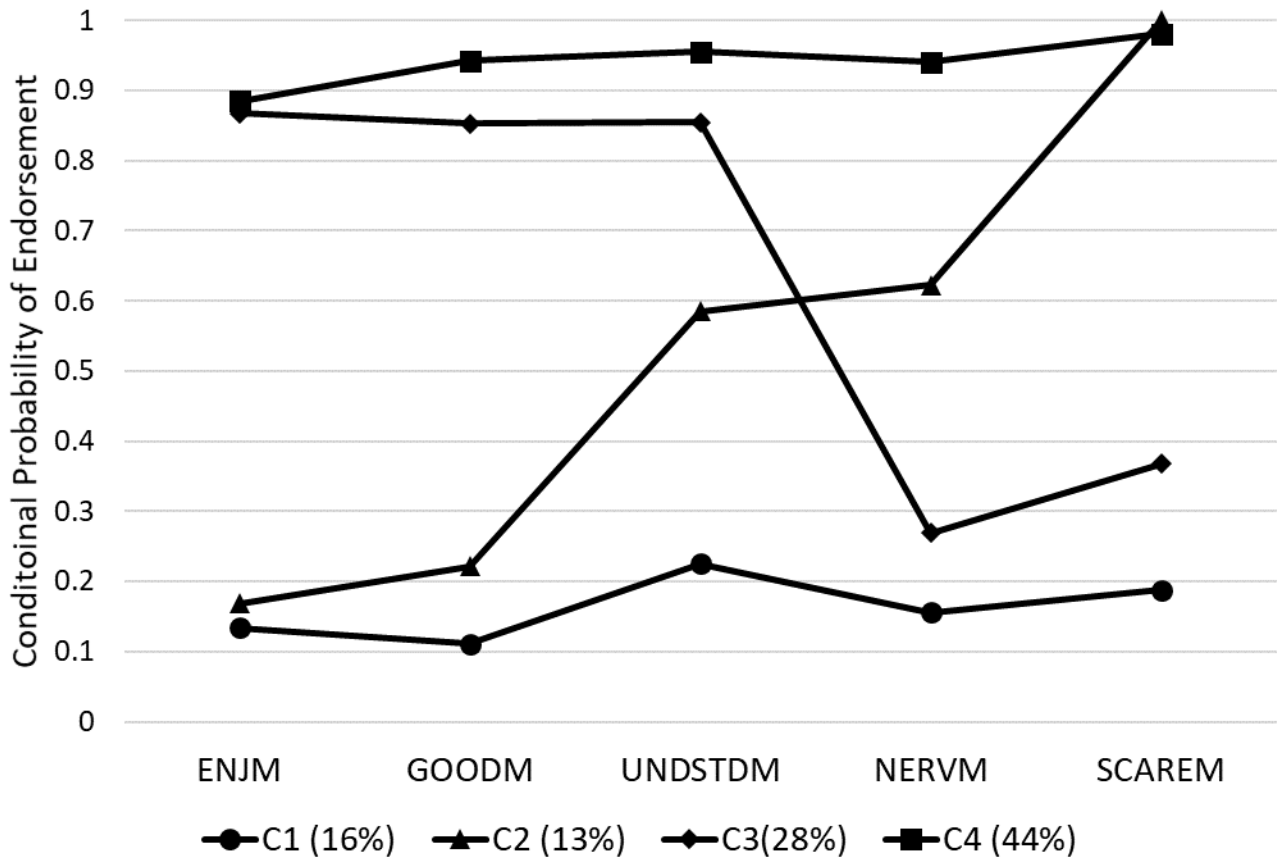
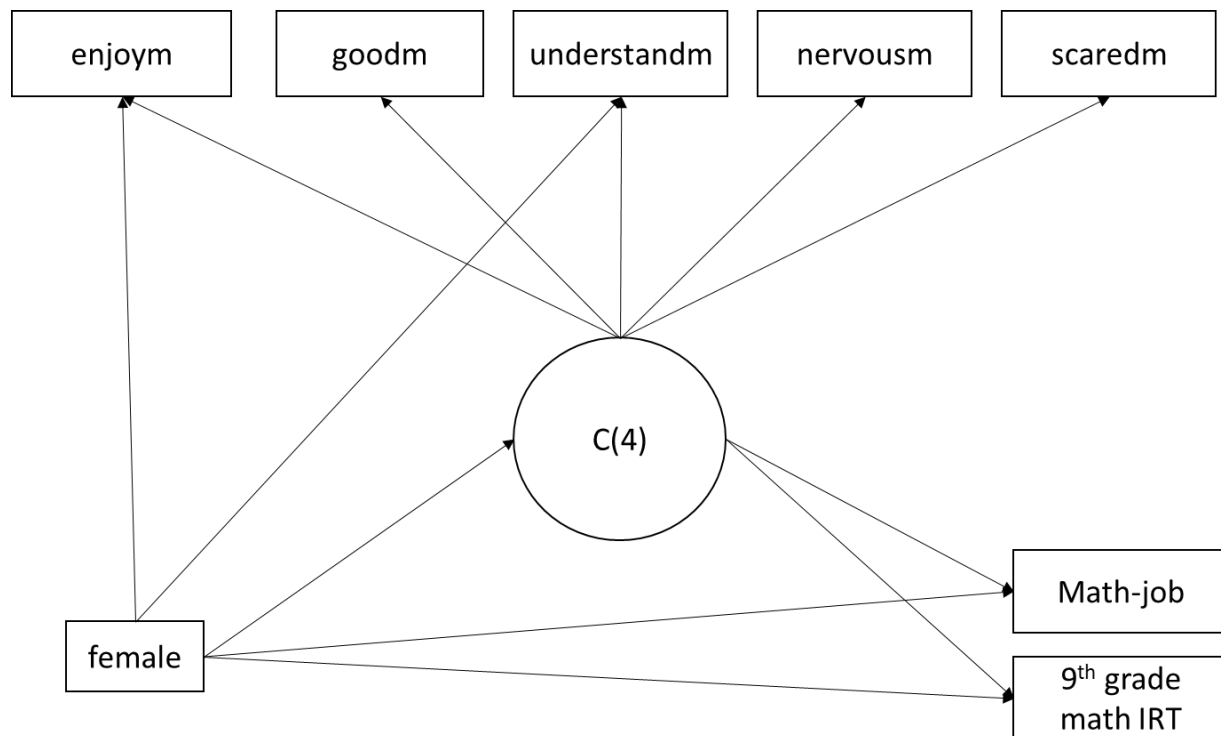


Figure 1. Item probability plots for the 4-class unconditional LCA plot





*Figure 2.* Model diagram of the 4-class LCA model with a covariate (female) and two distal outcomes (math-job and 9<sup>th</sup> grade math IRT), including DIF effects from the covariate to enjoy and understand.

## Appendix A: Distal-As- Indicator Model (Mplus Syntax)

### Variable:

```
usevariables = EA29KR emthirtn enjm goodm undstdm nervm scarem  
             female;  
categorical = EA29KR enjm goodm undstdm nervm scarem female;  
missing=all(9999);  
idvariable = lsayid;  
classes = c(4);
```

### Define:

```
!Renaming variables to make output easier to read  
enjm = ca28ar;  
goodm = ca28br;  
undstdm = ca28cr;  
nervm = ca28er;  
scarem = ca28gr;  
female = gender EQ 1;
```

### Analysis:

```
Type = mixture;  
starts = 100 50;  
processors = 4;
```

### Model:

```
%overall%  
  
%c#1%  
[EA29KR$1] (dj1);  
[emthirtn] (dm1);  
[female$1] (df1);  
  
%c#2%  
[EA29KR$1] (dj2);  
[emthirtn] (dm2);  
[female$1] (df2);  
  
%c#3%  
[EA29KR$1] (dj3);  
[emthirtn] (dm3);  
[female$1] (df3);  
  
%c#4%  
[EA29KR$1] (dj4);  
[emthirtn] (dm4);  
[female$1] (df4);
```

Output:

```
Tech11 Tech14;
```

!Following syntax is used to test the distal means

Model constraint:

```
New (dj1v2 dj1v3 dj1v4 dj2v3 dj2v4 dj3v4  
      dm1v2 dm1v3 dm1v4 dm2v3 dm2v4 dm3v4  
      df1v2 df1v3 df1v4 df2v3 df2v4 df3v4);
```

```
dj1v2 = dj1-dj2;    ! Creates a difference score comparing the  
                    distal mean for class 1 v class 2
```

```
dj1v3 = dj1-dj3;
```

```
dj1v4 = dj1-dj4;
```

```
dj2v3 = dj2-dj3;
```

```
dj2v4 = dj2-dj4;
```

```
dj3v4 = dj3-dj4;
```

```
dm1v2 = dm1-dm2;
```

```
dm1v3 = dm1-dm3;
```

```
dm1v4 = dm1-dm4;
```

```
dm2v3 = dm2-dm3;
```

```
dm2v4 = dm2-dm4;
```

```
dm3v4 = dm3-dm4;
```

```
df1v2 = df1-df2;
```

```
df1v3 = df1-df3;
```

```
df1v4 = df1-df4;
```

```
df2v3 = df2-df3;
```

```
df2v4 = df2-df4;
```

```
df3v4 = df3-df4;
```

## Appendix B: Manual 3-Step --- Step 1 (Mplus Syntax)

### Variable:

```
usevariables = enjm goodm undstdm nervm scarem;  
categorical = enjm goodm undstdm nervm scarem;  
auxiliary = gender raceth emthirtn ea29kr;
```

```
missing=all(9999);  
idvariable = lsayid;  
classes = c(4);
```

### Define:

```
!Renaming variables to make output easier to read  
enjm = ca28ar;  
goodm = ca28br;  
undstdm = ca28cr;  
nervm = ca28er;  
scarem = ca28gr;
```

### Analysis:

```
Type = mixture;  
starts = 0;  
processors = 4;
```

### Model:

```
%OVERALL%  
%C#1% !using start values to speed up estimation  
[ enjm$1*1.59328];  
[ goodm$1*1.25417];  
[ undstdm$1*-0.34487];  
[ nervm$1*-0.50261];  
[ scarem$1*-15];  
  
%C#2%  
[ enjm$1*-1.87899];  
[ goodm$1*-1.75626];  
[ undstdm$1*-1.77309];  
[ nervm$1*0.99921];  
[ scarem$1*0.54120];  
  
%C#3%  
[ enjm$1*-2.03944];  
[ goodm$1*-2.80241];  
[ undstdm$1*-3.06331];  
[ nervm$1*-2.77095];  
[ scarem$1*-3.94408];  
  
%C#4%
```

```
[ enjm$1*1.86433];  
[ goodm$1*2.07785];  
[ undstdm$1*1.23713];  
[ nervm$1*1.69216];  
[ scarem$1*1.46335];
```

Output: Tech1 svalues;

```
Plot: Type = plot3;  
series = enjm goodm undstdm nervm scarem (*);
```

```
SaveData:  
file is Step1_LSAY_4class_savedata_jobdistal.txt;  
save = cprob;  
format = free;  
missflag = 9999;
```

## Appendix C: Manual 3-step --- Step 3 (Mplus Syntax)

Variable:

```
Names are
  ENJM
  GOODM
  UNDSTDM
  NERVM
  SCAREM
  GENDER
  RACETH
  EMTHIRTN
  EA29KR
  CPROB1
  CPROB2
  CPROB3
  CPROB4
  CMOD
  LSAYID;
```

```
Missing are all(9999);
idvariable = lsayid;
usevariables = cmod emthirtn EA29KR female;! urm;
nominal = cmod;
classes = c(4);
```

Define:

```
female = gender EQ 1;
center female (grandmean);
```

Analysis:

```
type = mixture;
starts = 100 50;
processors = 4;
```

Model:

```
%overall%
c on female;           !latent class variable on cov
emthirtn on female;   !distal on cov
EA29KR on female;     !distal on cov
```

%c#1%

```
!syntax below is for the manual 3-step
[cmod#1@1.852  cmod#2@-0.095  cmod#3@0.310];
```

```
[emthirtn] (dml); !estimating class-specific mean for math
                !labeling it for testing later
```

```

    emthirtn;          !estimating class-specific var for math
    [EA29KR] (dj1);    !estimating class-specific math-job
    EA29KR;           !estimating class-specific var for math-job

%c#2%
    [cmod#1@-0.707    cmod#2@3.124    cmod#3@0.984];

    [emthirtn] (dm2);
    emthirtn;

    [EA29KR] (dj2);
    EA29KR;

%c#3%

    [cmod#1@3.617    cmod#2@5.406    cmod#3@7.903];

    [emthirtn] (dm3);
    emthirtn;

    [EA29KR] (dj3);
    EA29KR;

%c#4%

    [cmod#1@-2.348    cmod#2@-2.563    cmod#3@-5.954];

    [emthirtn] (dm4);
    emthirtn;

    [EA29KR] (dj4);
    EA29KR;

Model Test:
    0 = dm1 - dm2;
    0 = dm1 - dm3;
    0 = dm1 - dm4;

!These model constraint commands are used to create and test
!pairwise differences of each distal mean across the four
!classes.
! dm1v2 = distal "math" difference comparing class 1 and 2
! dm1v3 = distal "math" difference comparing class 1 and 3
! dj1v2 = distal "math-job" difference comparing class 1 and 2

Model Constraint:
    New(dm1v2 dm1v3 dm1v4 dm2v3 dm2v4 dm3v4

```

`dj1v2 dj1v3 dj1v4 dj2v3 dj2v4 dj3v4);`

`dm1v2 = dm1 - dm2;`

`dm1v3 = dm1 - dm3;`

`dm1v4 = dm1 - dm4;`

`dm2v3 = dm2 - dm3;`

`dm2v4 = dm2 - dm4;`

`dm3v4 = dm3 - dm4;`

`dj1v2 = dj1 - dj2;`

`dj1v3 = dj1 - dj3;`

`dj1v4 = dj1 - dj4;`

`dj2v3 = dj2 - dj3;`

`dj2v4 = dj2 - dj4;`

`dj3v4 = dj3 - dj4;`



## Appendix D: BCH-- Step 1 (Mplus Syntax)

### Variable:

```
usevariables = enjm goodm undstdm nervm scarem;  
categorical = enjm goodm undstdm nervm scarem;  
auxiliary = gender raceth emthirtn ea29kr;  
missing=all(9999);  
idvariable = lsayid;  
classes = c(4);
```

### Define:

```
! renaming variables to make output easier to read  
enjm = ca28ar;  
goodm = ca28br;  
undstdm = ca28cr;  
nervm = ca28er;  
scarem = ca28gr;
```

### Analysis:

```
Type = mixture;  
starts = 0;  
processors = 4;
```

### Model:

```
%OVERALL%  
  
%C#1%  
[ enjm$1*1.59328];  
[ goodm$1*1.25417];  
[ undstdm$1*-0.34487];  
[ nervm$1*-0.50261];  
[ scarem$1*-15];  
  
%C#2%  
[ enjm$1*-1.87899];  
[ goodm$1*-1.75626];  
[ undstdm$1*-1.77309];  
[ nervm$1*0.99921];  
[ scarem$1*0.54120];  
  
%C#3%  
[ enjm$1*-2.03944];  
[ goodm$1*-2.80241];  
[ undstdm$1*-3.06331];  
[ nervm$1*-2.77095];  
[ scarem$1*-3.94408];
```

```
%C#4%  
[ enjm$1*1.86433];  
[ goodm$1*2.07785];  
[ undstdm$1*1.23713];  
[ nervm$1*1.69216];  
[ scarem$1*1.46335];
```

Output: Tech1 svalues;

```
Plot: Type = plot3;  
series = enjm goodm undstdm nervm scarem (*);
```

```
SaveData:  
file is Step1BCH_LSAY_4class_savedata_jobdistal.txt;  
save = bchweights;  
format = free;  
missflag = 9999;
```

## Appendix E: BCH--Step 3 (Mplus Syntax)

Variable:

```
Names are
  ENJM
  GOODM
  UNSTDM
  NERVM
  SCAREM
  GENDER
  RACETH
  EMTHIRTN
  EA29KR
  BCHW1
  BCHW2
  BCHW3
  BCHW4
  CPROB1
  CPROB2
  CPROB3
  CPROB4
  CMOD
  LSAYID;
```

```
Missing are all(9999);
idvariable = lsayid;
usevariables = bchw1-bchw4 EA29KR emthirtn female;! urm;
classes = c(4);
training = bchw1-bchw4 (bch);
auxiliary = cmod;
```

```
Define:
  female = gender EQ 1;
  center female (grandmean);
```

```
Analysis:
  type = mixture;
  starts = 100 50;
  processors = 4;
```

```
Model:
  %overall%
  c on female;
  emthirtn on female;
  EA29KR on female;
```

%c#1%  
[emthirtn] (dm1);  
emthirtn;  
  
[EA29KR] (dj1);  
EA29KR;  
  
emthirtn on female (rm1);  
EA29KR on female (rj1);

%c#2%  
[emthirtn] (dm2);  
emthirtn;  
  
[EA29KR] (dj2);  
EA29KR;  
  
emthirtn on female (rm2);  
EA29KR on female (rj2);

%c#3%  
[emthirtn] (dm3);  
emthirtn;  
  
[EA29KR] (dj3);  
EA29KR;  
  
emthirtn on female (rm3);  
EA29KR on female (rj3);

%c#4%  
[emthirtn] (dm4);  
emthirtn;  
  
[EA29KR] (dj4);  
EA29KR;  
  
emthirtn on female (rm4);  
EA29KR on female (rj4);

Model Test:

0 = dm1 - dm2;  
0 = dm1 - dm3;  
0 = dm1 - dm4;

Model Constraint:

```
New(dm1v2 dm1v3 dm1v4 dm2v3 dm2v4 dm3v4  
dj1v2 dj1v3 dj1v4 dj2v3 dj2v4 dj3v4);
```

```
dm1v2 = dm1 - dm2;  
dm1v3 = dm1 - dm3;  
dm1v4 = dm1 - dm4;  
dm2v3 = dm2 - dm3;  
dm2v4 = dm2 - dm4;  
dm3v4 = dm3 - dm4;
```

```
dj1v2 = dj1 - dj2;  
dj1v3 = dj1 - dj3;  
dj1v4 = dj1 - dj4;  
dj2v3 = dj2 - dj3;  
dj2v4 = dj2 - dj4;  
dj3v4 = dj3 - dj4;
```

```
Output: Tech1 svalues sampstat;
```

```
SaveData:
```

```
file is Step3BCH_LSAY_4class_interact_savedata_jobdistal.txt;  
save = cprob;  
format = free;  
missflag = 9999;
```

## Appendix F: Relevant Syntax Specifying Covariate and Distal Outcome Effects Using the ML Approach After Accounting for DIF Effects (Latent Gold Syntax)

```
options
  !GUI options left out for brevity
  step3 modal ml;

variables
  independent GENDER;
  dependent math-job nominal, math-IRT continuous;
  latent Cluster nominal posterior = (clu#1 clu#2 clu#3 clu#4);

equations
  Cluster <- 1 + GENDER;
  math-job <- 1 + Cluster;
  math-IRT <- 1 + Cluster;
  math-IRT;
```

## Appendix G: Relevant Syntax Specifying Covariate and Distal Outcome Effects Using the BCH Approach After Accounting for DIF Effects (Latnet Gold syntax)

```
options
  !GUI options left out for brevity
  step3 modal bch;

variables
  independent GENDER;
  dependent math-job nominal, math-IRT continuous;
  latent Cluster nominal posterior = (clu#1 clu#2 clu#3 clu#4);

equations
  Cluster <- 1 + GENDER;
  math-job <- 1 + Cluster;
  math-IRT <- 1 + Cluster;
  math-IRT;
```