Technical Report of Methodological Critique of Swartout et al., 2015

The July 13, 2015 JAMA article entitled, "Trajectory Analysis of the Campus Serial Rapist Assumption" authored by Swartout, Koss, White, Thompson, Abbey, and Bellis reports a study conducted with sizeable samples from both a derivation and validation dataset to fit latent trajectories of the probability of raping across the college years, using perpetrators' own reports. Based on the results of their study, the authors conclude that most campus rapists are *not* predatory serial rapists but rather opportunistic, time-limited rapists.

This technical report uses the original Swartout et al. analysis dataset and syntax, alterations to the syntax to test various model assumptions and an examination of the public-use dataset that was used to create the article's derivation dataset.

JAMA Article Latent Trajectory Models

Derivation dataset: 5-timepoint model (Appendix A). Despite the authors' report that the fifth timepoint was omitted from the derivation analysis due to low response rates (p. E3, 1st paragraph), the Mplus syntax file provided by Swartout included the fifth timepoint. The inclusion of the fifth timepoint was not an oversight in the version of the syntax file Swartout provided, since the output matches the model fit statistics given in the article (although not perfectly) and the plots produced by the syntax match the graphs included in the article's figure. Further, when the fifth timepoint is dropped (see section below), the results do not replicate the values and graph reported in the article.

In these data, the covariance coverage (proportion of data present) is as low as 11% across the sophomore and senior timepoints and as low as 17% within the senior timepoint. The consequence of the use of the "low response" fifth wave in the derivation dataset is that the analysis relies heavily on the missing data algorithm and the assumption of missing at random on which it is based. The authors claim (E3, 1st paragraph) that "…missing data were not related to reports of sexual violence across the study" reporting a non-significant Pearson chi-square test for missing completely at random (MCAR). When the MCAR assumption holds, the missingness is "ignorable" (i.e., does not produce biased estimates). In this case, there are no covariates in the analysis so the test of MCAR also tests the assumption of missing at random (MAR) – in other words, the assumption that missingness in the dataset can be explained by variables included in the analysis model.

The analysis produced warning that the information matrix was singular (the model was empirically under-identified). In this case, the singular matrix is likely to be due to empty cells in the multiway contingency table as a result of having few reported rapes (5%, 3%, 3%, 4%, and 1%, for each successive timepoint) and a great deal of missing data. This singularity required two parameters to be fixed - the slope and quadratic growth parameters for the Increasing class, central parameters of interest. The choice of which parameter(s) to fix and at what values was not researcher-driven but instead automatically selected by the Mplus software.

If we assume that fixed values for the slope and quadratic parameters are reasonable, we can evaluate model fit statistics. Perhaps the strongest empirically is the likelihood ratio test (LRT), which compares the fit of the model run with the fit of a model with one fewer classes. Swartout's syntax requests 3 versions of the LRT: the original Vuong-Lo-Mendell-Rubin LRT, the Lo-Mendell-Rubin adjusted LRT, and the parametric bootstrapped LRT. Two of these indices are statistically significant (the current model improves the model with one fewer classes) but the bootstrapped LRT is not significant.

There are no standardized fit statistics such as CFI available for mixture models. Instead, successive mixture solutions can be evaluated through comparative indexes such as the adjusted BIC, class size, and classification quality. As reported in the article, the adjusted BIC for the 3-class solution is between those for the 2-class and 4-class solution, suggesting that a maximal value had not been reached with 4 classes. I assume that Swartout selected the 3-class solution over the others based on the LRT and/or the entropy value, since these two descriptors point to the 3-class solution. However, the class sizes (ideally, greater than 5% and at least 30 members) and classification quality were not optimal for the 3-class solution. An estimate based on calculated probabilities yields (within rounding error) 16 (2%) in the Increasing class, 822 (97%) in the Low or Time-limited class, and 12 (1%) in the Decreasing class. However, when individuals are assigned to the class for which their probability highest, the sizes are 12 (1%), 830 (98%) and 8 (1%), respectively. Both estimates show that two of the three class sizes are extremely small.

The overall classification quality is given by the entropy value for this model, .939. Mplus also produces two classification matrices against which the local fit (classification quality specific to each class) can be evaluated. The first compares the relative probabilities within individuals and the second compares the relative probabilities within classes. In the output generated by Swartout's syntax, estimated counts and proportions are quite different across the two matrices and both reveal very poor classification for the two smallest classes. Individuals assigned to the Increasing latent class have only a .675 probability of belonging to that class but have .324 probability of belonging to the Low or Time-limited class. A similar story is apparent in the second table; within the Increasing class, the probability of "correct" classification is .511 whereas the probability of being "misclassified" into the Low or Time-limited class is .489. Among individuals assigned to the Decreasing class, the average probability of belonging to that class is .954 (1st table) but the probability of "correct" classification is .612 and the probability of being "misclassified" to the Low or Time-limited class is .387 (2nd table).

If we ignore the small class sizes and classification discrepancies and if we assume that the 3-class solution is indeed optimal, we can evaluate the model parameters. Since the syntax specified that the trajectory was centered on R1, the intercept represents raping prior to college and the slope roughly represents the slope from pre-college to freshman year. Being the highest order trajectory parameter, the quadratic term describes the rate of deceleration in this slope across the entire study period. The plot requested in Swartout's syntax file is included with the output in Appendix A.

The estimated intercept for the Increasing class was -59.984 (p<.001). Recall that the slope and quadratic terms were fixed within the model estimation at 79.859 and -17.574, respectively. The scale of these mean parameters is much larger than the scale of the parameters of the other two classes. In the Decreasing class, the estimated intercept is 8.837 (p<.001), the slope is -4.194 (p<.05), and the quadratic is .624 (*ns*). The mean for the Low or Time-limited class is set to zero by default in order to identify the scale of the trajectory means. The slope parameter was estimated to be -.866 (p<.05) for this class and the quadratic term was .156 (*ns*). Despite the extreme differences across groups in trajectory parameters, odds ratio tests were not significant.

In the original syntax file, several lines of code requesting bootstrapped standard errors were commented out. Bootstrapped standard errors do not require the assumption of normal sampling distributions of the estimates. This is a reasonable choice for this application because the parameters are at the extremes of admissible values (0 and 1 due to the probability scale). However, the Increasing and Decreasing classes are so sparse that there is too little information from which to make bootstrapped draws, resulting in extremely inflated rather than tighter standard errors. Because of the inflated standard errors, none of the estimated trajectory parameters were significant in the bootstrapped model.

Derivation dataset: Monte Carlo power simulation (Appendix B). Using the parameters generated by Swartout's derivation syntax, a simulation study reasonably reproduced the population values given for the key parameters. In this simulation, the statistical power associated with the trajectory parameters was high for the Increasing and Decreasing classes and (appropriately) low for the Low or Time-limited class. Power was also good for distinguishing between the Increasing and Low or Time-Limited trajectory classes and between the Decreasing and Low or Time-limited classes. However, the simulation produced serious warnings for each of the 240 replications that indicate that the model is under-identified and required a good number of parameters to be fixed in order to estimate the model.

Derivation dataset: 4-timepoint model (Appendix C). When the final timepoint is *excluded* from the derivation analysis model, there is a warning of under-identification where a number of critical trajectory parameters (slopes and quadratic terms for the Increasing and Decreasing classes) are fixed by the Mplus program in order to identify the model. The missing data covariance coverage was as low as .182 and the Decreasing and Increasing classes were very small (n=8 & 6). The test for MCAR/MAR is significant and must be rejected. This makes the missing data handling technique used in the JAMA article very problematic, especially given the extent of missing data across time.

While the entropy value is reasonable (0.951), there is evidence of strong misclassification; based on the estimated probabilities, 32% of those assigned to the Decreasing class were misclassified as the Low or Time-limited class and 65% of those assigned to the Increasing class were misclassified as the Low or

Time-limited class. In the 4 timepoint model, estimated probabilities of raping at 3 of the 4 timepoints were 0 or 1. Patterns that yield probabilities at the boundary of admissible values are often a result of too few members in a given class, resulting in parameter values that are overfit to the data from these individuals. In fact, for the Increasing and Decreasing classes, the model produced improbably high values for both fixed and estimated parameters. Because the standard errors are so large, the two classes are not statistically distinguishable from one another, even though the trajectory parameters are very large and in the opposite directions. The graph of probability estimates calculated from the parameter estimates showed probabilities similar to the 5-timepoint model in the Increasing and Low or Time-limited classes but what might have been the Decreasing class has a predicted probability of raping of 1.00 at 3 of the 4 timepoints.

Validation dataset (Appendix D). The model using the validation dataset had fewer missing data, which led to slightly higher estimates of class membership and model identification (no parameters were fixed to avoid singularity. When the trajectory parameters are estimated rather than fixed, it is apparent that all but 2 of the critical trajectory estimates are non-significant and the Decreasing class is not distinguishable from the Increasing class. The results of the model are similar to but do not exactly match those given in the JAMA article.

Validation dataset: Monte Carlo power simulation (Appendix E). In this power simulation, the power associated with all trajectory parameters for all classes were < .60 and the power for distinguishing between the Increasing and Decreasing trajectory classes was .44.

Alternative Models Using Swartout's Data

Subsample of rapists only: Latent trajectory analysis (Appendix F). It is not unusual for a very large subpopulation, such as the non-rapists included in these datasets, to heavily influence the latent class solution. To focus the analysis on the population of interest, namely campus rapists, I omitted the non-rapists (no rape at any timepoint) and combined the derivation and validation datasets to increase the number of rapists on which to base a trajectory analysis. Using the same syntax provided by Swartout, I arrived at a very different story. Like the JAMA article models, the best fitting model using the parametric bootstrapped LRT pointed to the 3-class solution (alternative solutions not shown). In this model, the proportion of the sample admitting to rape at each timepoint is substantially greater and, because the two datasets are combined, the data are somewhat less sparse but the test of MCAR/MAR is significant so the missing data assumption is not supported. Further, the estimation problems seen in the JAMA articles (e.g., model under-identification) continue to be a problem.

In this model, the latent classes are more reasonably balanced (61%, 22%, & 17%) but there is evidence of overall and class-specific misclassification.

Only the smallest class has a pattern that describes time-limited rape during the college years (freshman year) but this conclusion is complicated by the fact that all three of the trajectory parameters were fixed. Of the other two classes, only the class with generally increasing predicted probabilities had significant trajectory parameters whereas the class represented by monotonically declining predicted probabilities had no significant trajectory estimates. The time-limited and decreasing classes were not distinguishable from one another. However even if these classes were combined, the increasing latent class would still represent 50% more of the sample (n=49 & 78, respectively).

Subsample of rapists only: Latent profile analysis (Appendix G). By using trajectory analysis, the authors impose a restriction on the data that the probability of rape can best be described by a smooth line over time – a strict constraint that precludes men who rape at non-consecutive timepoints from being identified as serial rapists. Further, using the pre-college timepoint for which all participants have data heavily influences the growth trajectory estimates, particularly since the centering of the trajectory on this timepoint completely describes the intercept term and defines the slope which is estimated at the pre-college timepoint as the tangent to the curved line. Indeed, the “increasing” class in the JAMA article has a 0% probability of rape at pre-college and the “decreasing” class has a 100% probability of rape at pre-college. Such heavy reliance on the pre-college timepoint, paired with the implied independence of pre-college and college rapes (E6, 1st paragraph), does not directly address rapes occurring *during the college years*.

As an alternative model, I relaxed the constraint of a smooth growth curve across time and obtained a 5-class model (model fit comparisons not shown). Despite more liberal model assumptions, this model produced the same warning of model under-identification. Optimization that produces parameters near the boundary of admissible values, namely the threshold parameters on which the predicated probabilities are based, resulted in thresholds fixed at values producing 1 or 0 probabilities. As pointed out above, this is not uncommon when the sample sizes for the latent classes are small. In this model, the smallest classes had an estimated membership of only 12 and 14 individuals and the classification certainty for these two classes was lower than for the other classes. In addition, none of the classes were statistically distinguishable from the arbitrarily assigned final class and the tests of comparisons at specific timepoints were largely inestimable. An examination of the graph of predicted probabilities for each class shows that 3 of the 5 classes (65% of the rapist sample) had 40% or higher probability of rape for at least two of the four timepoints.

Autoregression: Derivation dataset (Appendix H). Rather than fitting a complicated and computationally intensive latent class model, I fit a simpler alternative model in which one timepoint is regressed on rape at the immediately previous timepoint (logistic autoregression). This model converged without difficulty and minimized the impact of missing data since estimates were obtained only for consecutive timepoints. This model had excellent fit and each regression parameter was significant to the $p<.001$ level. Based on this model, the predicted probability of raping during freshman year was nearly 4 times higher (OR=3.79) if a man had raped pre-college; the probability of raping during sophomore year was more than twice as high (OR=2.34) if a man had raped during freshman year. Although the parameter estimate narrowly missed the $p<.05$ level due to missing data, the probability of raping during junior year was more than 2 times higher (OR=2.76) if a man had raped during sophomore year. While the effects of raping at an earlier timepoint are cumulative, these indirect effects on the junior year, while sizable, are not statistically significant due to missing data. The significant indirect effect on sophomore year represents an accumulated risk: a higher likelihood of raping again as a freshman based on a pre-college rape further increases the chance of raping during the sophomore year over those who raped during freshman year but not pre-college.

Autoregression: Validation dataset (Appendix I). When the validation dataset is used, the autoregression model converges quickly, has very good model fit statistics, and results in significant and sizable estimates for all direct and indirect effects from pre-college through the senior year. Predicted probabilities based on this model graphically show the model results, where the likelihood of raping again during freshman year is nearly 7 times higher if a man raped pre-college. The likelihood of raping again during sophomore year is twice as high when a man reports raping during freshman year. The odds of raping again during junior year are 2 ½ times higher if a man raped during sophomore year. The odds of raping again during senior year are twice as high if a man reported raping during junior year.

Reconstruction of Analysis Variables in the Original Data (Appendix J)

I directly examined the original public-use dataset from which Swartout et al created their derivation dataset, the *Longitudinal Study of Violence Against Women: Victimization and Perpetration Among College Students in a State-Supported University in the United States, 1990-1995* (ICPSR 3212: http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/3212). Unfortunately, the original case ID numbers were absent in the Mplus analysis dataset used for the article and (despite repeated requests) no file was provided to match the JAMA dataset to the public-use dataset so an examination of differential missingness and demographic and other characteristics of the members of the latent trajectory classes was not possible. Since missing data have been shown to be problematic in the analysis, the extent of missing data, the unexplained missing data patterns, and the missing data miscoded as no rape are serious issues challenging the validity of the analysis and conclusions reported in the JAMA article. To examine the sources of missingness and how missing data impacted the construction of the analysis variables, I used information available to me through the description in the JAMA article, the SPSS syntax code provided by Swartout, and the documentation associated with the public-use dataset.

Using the SPSS syntax code exactly as it was provided by Swartout, I constructed a dichotomous rape variable for each timepoint. Frequencies of the reconstructed variables did not match the frequencies of the rape variables in the JAMA analysis dataset. Swartout's dataset and reconstruction in the public-use dataset identified rapists for the following timepoints: 44 for both in the pre-college timepoint, 20 for both in the freshman year, 11 in the JAMA dataset and 12 in the public dataset for sophomore year, 10 in the JAMA dataset and 12 in the public dataset for junior year, and 1 in both datasets for senior year. In addition, the extent of missing data in the reconstructed variables did not match the missing data in the JAMA analysis dataset. In the pre-college timepoint, the JAMA dataset had 4 missing cases and the public dataset had none. For freshman year, JAMA had 215 and the public data had 204 missing cases. For sophomore year, the missing cases were 507 for JAMA and 395 for the public data. In the junior year, JAMA had 624 missing cases and the public data had 551. Both datasets had 706 missing cases for the senior year.

Oddly, Swartout's code did not draw directly from the raw frequency data that the respondents provided. Instead, Swartout's syntax drew from a pre-constructed composite of sexual experiences per timepoint (expgrp#) which was in turn based on a dichotomization of each of three frequency-based sexual assault items. Importantly, a cross-tabulation of the dichotomized sexual assault indicators with the frequency data shows that, in some cases, missing data were assigned as "no rape": 37 in pre-college, 20 in freshman year, 17 in sophomore year, and 11 in junior year. None were miscoded in senior year.

In an attempt to understand this disconnect, I tried to reconstruct the sexual experiences variable used in Swartout's syntax. This exercise revealed that 395 of the study participants who had dropped out by junior year had "never" codes for that year on two variables on which the sexual experience composite was based. I also corrected miscoding in a key sexual assault variable (DRUGSI5) where codes of "never" (1) were mistakenly assigned as missing for senior year, resulting in 850 missing cases. These mistakes appear to have been correct in the creation of the expgrp# variables (although incorrect in the publically available dataset) because most assignments matched across these two versions. However, a few cases had conflicting assignments across the original and reconstructed versions of expgrp#: one in pre-college, 8 in freshman year, 4 in sophomore year, and 2 in junior year. There were no mismatched assignments in senior year.

In addition to study attrition, there appears to be a substantial source of unexplained missingness where men who did participate in the study for a given timepoint did not provide data for any of the sexual assault indicators. A small amount of missingness resulted from some cases not following the logic of the construction of the sexual experience categories and therefore being assigned to none of the categories. Further, spot missingness (one or two of the three sexual assault items used to operationalize rape) was not appropriately dealt with when constructing the dichotomous rape variable used for analysis. If any of the three sexual assault indicators is endorsed, then "rape" can be assigned for that timepoint. If all indicators are denied, then "no rape" can be assigned. However, if there is a mix of missing and denial, true rape status is not known (missing values could potentially mask an endorsement) and since rape status is uncertain, the dichotomous rape variable must be considered missing.

In the end, I constructed a rape indicator per timepoint that drew directly from the frequency variables and used correct logic to assign missingness, "no rape" and "rape." A cross-tabulation of these rape indicators and those created by simply running Swartout's code based on the original expgrp#

variables shows mismatched cases: 19 pre-college, 19 freshman year, 14 sophomore year, and 5 junior year.

Operationalization of Rape

It is unfortunate that Swartout and his colleagues chose not to use the full information about sexual assault available in the dataset. By creating a single dichotomous indicator of rape for each timepoint, the authors are ignoring multiple rapes within timepoint, underestimating serial rapists. Further, by assigning a single rape indicator to multiple reports of rape across sexual assault items, the authors are assuming all assault indicators occurred within the same rape incident, underestimating serial rapists. By ignoring rape attempts, the authors are underestimating serial rapists. Even as an academic outside this field of expertise, these seem important considerations.

Summary

Based on my examination of the data and analysis models used in the JAMA article, I conclude that the scientific integrity of the study and, by extension, the conclusions based on the study are highly suspect at best. As it stands, the article relies on erroneously coded data, misalignment with the raw data, and untenable model assumptions, including the assumptions underlying the method used to handle a large amount of missing data. It is my opinion that reasonable debate over the serial campus rapist assumption cannot ride on this study.

Appendix A: Annotated Output Generated by Swartout's Syntax Files – Derivation Dataset

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
09/20/2015   2:57 PM

INPUT INSTRUCTIONS

  TITLE: Final runs of R Traj analysis - Derivation Data;
  DATA: FILE IS RTraj.dat;
  VARIABLE: NAMES ARE id R1 R2 R3 R4 R5 dataset use;
  useobs is (dataset eq 3);
  IDVARIABLE = id;
  USEVAR = R1-R5;
  categorical ARE R1-R5;
  MISSING ARE all(9999999);
  CLASSES =  c(3);
  ANALYSIS: TYPE = MIXTURE;
  Estimator = MLR;
  !!! The following commands for bootstrapping were in the original file, commented out.
  !for bootstrapped confidence intervals;
  !Estimator = ML;
  !BOOTSTRAP = 15000;
  Processors = 7;
  starts= 1000 200;
  MODEL:
  %OVERALL%
  i s q| R1@0 R2@1 R3@2 R4@3 R5@4;
  !!! Note that 5 timepoints are used for the trajectory
  OUTPUT: tech11 tech14 CINTERVAL(BCBOOTSTRAP) ;
  PLOT:
  type = plot3;
  series = R1-R4 (s);
  !!!! Note that 4 timepoints are used for the JAMA graph

*** WARNING
  Data set contains cases with missing on all variables.
  These cases were not included in the analysis.
  Number of cases with missing on all variables:  1
   2 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS

Final runs of R Traj analysis - Derivation Data;
SUMMARY OF ANALYSIS
Number of observations                                          850
SUMMARY OF DATA
     Number of missing data patterns            14
     Number of y missing data patterns           0
     Number of u missing data patterns          14

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100
     PROPORTION OF DATA PRESENT FOR U
           Covariance Coverage
              R1            R2            R3            R4            R5
              ________      ________      ________      ________      ________
 R1             0.996
 R2             0.745         0.748
 R3             0.404         0.392         0.405
 R4             0.265         0.261         0.182         0.267
 R5             0.168         0.171         0.108         0.133         0.171
 !!! Note the extent of missing data across time (off diagonals) and within time (diagonals)
UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES
    R1
      Category 1    0.948         803.000
      Category 2    0.052          44.000
    R2
      Category 1    0.969         616.000
      Category 2    0.031          20.000
    R3
      Category 1    0.968         333.000
      Category 2    0.032          11.000
    R4
      Category 1    0.956         217.000
      Category 2    0.044          10.000
    R5
      Category 1    0.993         144.000
      Category 2    0.007           1.000
      !!! Note sparseness of rape category within and across timepoints
      !!!(.7% to 5.2%, n=1 to 44)

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES
Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
            -335.851  292884           103
```

```
           -335.851  416250           390
           -335.851  863691           481
           -335.851  572637           989
           .
           .
           .
           -337.588  845580           805
           -337.588  165853           105
           -337.588  468036           131
           -337.588  923437           398
           -337.588  752769           253
           -337.588  82200            830
           -337.588  714997           399
           -337.588  824126           287

THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED.  RERUN WITH AT LEAST TWICE THE
RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.
     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
!!! The joint distribution is a problem due to the sparseness of the
!!! rape category and missingness within and across timepoints.
     THE FOLLOWING PARAMETERS WERE FIXED:
     Parameter 4, %C#1%: [ Q ]
     Parameter 3, %C#1%: [ S ]
!!! Growth parameters for the Increasing class were fixed, not estimated.

THE MODEL ESTIMATION TERMINATED NORMALLY
MODEL FIT INFORMATION
Number of Free Parameters                        11
Loglikelihood
          H0 Value                          -335.851
          H0 Scaling Correction Factor        0.6051
            for MLR
Information Criteria
          Akaike (AIC)                       693.701
          Bayesian (BIC)                     745.899
          Sample-Size Adjusted BIC           710.966
            (n* = (n + 2) / 24)
            !!! The adjusted BIC does not match that reported in the article
            !!! BLRT is reported at bottom of output file
Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes
          Pearson Chi-Square
          Value                               17.945
          Degrees of Freedom                      20
          P-Value                             0.5910
          Likelihood Ratio Chi-Square
          Value                               11.398
          Degrees of Freedom                      20
          P-Value                             0.9352
Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator Model
          Pearson Chi-Square
          Value                               66.038
          Degrees of Freedom                     103
          P-Value                             0.9983
          Likelihood Ratio Chi-Square
          Value                               41.835
          Degrees of Freedom                     103
          P-Value                             1.0000
!!! Supports the MCAR assumption.

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL
    Latent
   Classes
       1        15.84924          0.01865 !!! Increasing class
       2        12.45556          0.01465 !!! Decreasing class
       3       821.69519          0.96670 !!! Low or Time-limited class

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES
    Latent
   Classes
       1        15.84925          0.01865
       2        12.45556          0.01465 !!! This value matches that reported in the article
       3       821.69519          0.96670 !!! to 2 decimal places but, due to misclassification
                                          !!! (see criterion below), evidence for this being the
                                          !!! smallest class is mixed.
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP
Class Counts and Proportions
    Latent
```

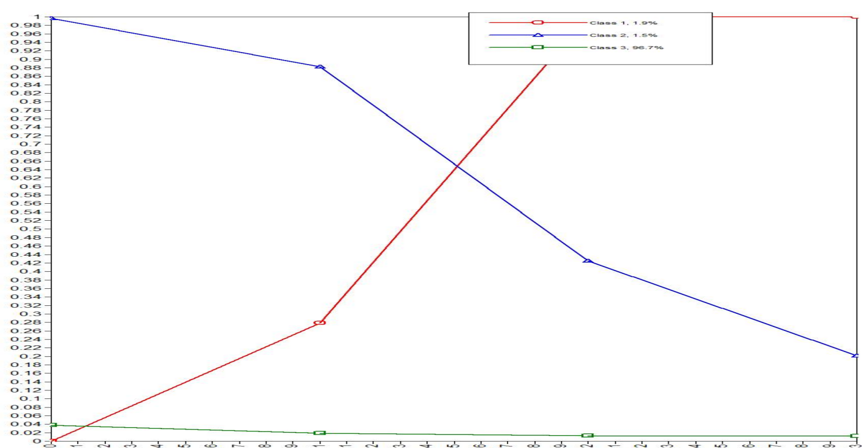
```
   Classes
       1               12          0.01412
       2                8          0.00941
       3              830          0.97647
CLASSIFICATION QUALITY
     Entropy                         0.939
     !!! This value matches that reported in the article to 2 decimal places.

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)
           1        2        3
    1   0.675    0.001    0.324 !!! Class 1 ("Increasing" class) is cross-
    2   0.000    0.954    0.046 !!! classified with Class 3 ("Low or Time-limited" class).
    3   0.009    0.006    0.985 !!! These values nearly match those reported in
                                !!! the article.
Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
           1        2        3
    1   0.511    0.000    0.489 !!! Here, the misclassification is even more evident.
    2   0.001    0.612    0.387
    3   0.005    0.000    0.995
Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
              1        2        3
    1      0.044  -13.100    0.000
    2     -6.494    0.459    0.000
    3     -5.347   -7.695    0.000

MODEL RESULTS
!!! None of the quadratic terms in this model are significant. A subsequent run (not shown)
!!! that omitted the quadratic term resulted in results with the same problems as noted below
!!! and none of the slope parameters were statistically significant.
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Latent Class 1 !!! Increasing class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                -59.984      0.950    -63.118      0.000
    S                 79.859      0.000    999.000    999.000 !!! These two parameters were fixed
    Q                -17.574      0.000    999.000    999.000 !!! within the run to avoid
                                                              !!! singularity.
 Thresholds
    R1$1               3.251      0.213     15.250      0.000
    R2$1               3.251      0.213     15.250      0.000
    R3$1               3.251      0.213     15.250      0.000
    R4$1               3.251      0.213     15.250      0.000
    R5$1               3.251      0.213     15.250      0.000
 !!! These lines were taken from the bootstrapped run, which is designed to decrease,
 !!! rather than increase the standard errors.
 !!! Means
 !!!   I                -61.056  12475.554     -0.005      0.996
 !!!   S                 81.263  11251.050      0.007      0.994
 !!!   Q                -17.906   4605.808     -0.004      0.997
 !!! Thresholds
 !!!   R1$1               3.251   1223.907      0.003      0.998
 !!!   R2$1               3.251   1223.907      0.003      0.998
 !!!   R3$1               3.251   1223.907      0.003      0.998
 !!!   R4$1               3.251   1223.907      0.003      0.998
 !!!   R5$1               3.251   1223.907      0.003      0.998

Latent Class 2 !!! Decreasing class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
```

```
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                  8.837      2.098      4.212      0.000
    S                 -4.194      2.003     -2.094      0.036
    Q                  0.624      0.391      1.598      0.110 !!! Non-significant quadratic term
 Thresholds
    R1$1               3.251      0.213     15.250      0.000
    R2$1               3.251      0.213     15.250      0.000
    R3$1               3.251      0.213     15.250      0.000
    R4$1               3.251      0.213     15.250      0.000
    R5$1               3.251      0.213     15.250      0.000
 !!! These lines were taken from the bootstrapped run.
 !!! Means
 !!!    I                  8.837  *********      0.000      1.000
 !!!    S                 -4.194  *********      0.000      1.000
 !!!    Q                  0.624  54133.633      0.000      1.000
 !!! Thresholds
 !!!    R1$1               3.251   1223.907      0.003      0.998
 !!!    R2$1               3.251   1223.907      0.003      0.998
 !!!    R3$1               3.251   1223.907      0.003      0.998
 !!!    R4$1               3.251   1223.907      0.003      0.998
 !!!    R5$1               3.251   1223.907      0.003      0.998

Latent Class 3 !!! Low or Time-limited class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000

 Means
    I                  0.000      0.000    999.000    999.000
    S                 -0.866      0.414     -2.092      0.036
    Q                  0.156      0.104      1.504      0.133 !!! Non-significant quadratic term

 Thresholds
    R1$1               3.251      0.213     15.250      0.000
    R2$1               3.251      0.213     15.250      0.000
    R3$1               3.251      0.213     15.250      0.000
    R4$1               3.251      0.213     15.250      0.000
    R5$1               3.251      0.213     15.250      0.000

Categorical Latent Variables
 Means
    C#1               -3.948      0.463     -8.534      0.000
    C#2               -4.189      0.375    -11.185      0.000

RESULTS IN PROBABILITY SCALE
Latent Class 1
 R1
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000 !!! Parameters that approach the
 R2                                                           !!! limit (at or near 100% or 0%
    Category 1         0.721      0.183      3.937      0.000 !!! probability) produce
    Category 2         0.279      0.183      1.523      0.128 !!! computational problems and
 R3                                                           !!! potentially biased standard errors.
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
 R4
    Category 1         0.000      0.000      0.000      1.000
```

```
    Category 2         1.000      0.000      0.000      1.000
 R5
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000 !!! This is the value for the
                                                              !!! R5 timepoint not shown in the article.
Latent Class 2
 R1
    Category 1         0.004      0.008      0.475      0.635
    Category 2         0.996      0.008    126.656      0.000
 R2
    Category 1         0.117      0.093      1.256      0.209
    Category 2         0.883      0.093      9.440      0.000
 R3
    Category 1         0.575      0.214      2.687      0.007
    Category 2         0.425      0.214      1.983      0.047
 R4
    Category 1         0.798      0.119      6.731      0.000
    Category 2         0.202      0.119      1.701      0.089
 R5
    Category 1         0.768      0.018     41.880      0.000
    Category 2         0.232      0.018     12.628      0.000
Latent Class 3
 R1
    Category 1         0.963      0.008    125.772      0.000
    Category 2         0.037      0.008      4.873      0.000
 R2
    Category 1         0.981      0.005    212.868      0.000
    Category 2         0.019      0.005      4.057      0.000
 R3
    Category 1         0.987      0.005    218.486      0.000
    Category 2         0.013      0.005      2.802      0.005
 R4
    Category 1         0.988      0.005    218.106      0.000
    Category 2         0.012      0.005      2.573      0.010
 R5
    Category 1         0.985      0.008    124.611      0.000
    Category 2         0.015      0.008      1.849      0.064
LATENT CLASS ODDS RATIO RESULTS
Latent Class 1 Compared to Latent Class 2
 R1
    Category > 1       0.000      0.000      0.433      0.665 !!! None of these tests of
 R2
    Category > 1       0.051      0.066      0.774      0.439 !!! differences in odds ratios across
 R3
    Category > 1   *********  *********      0.787      0.432 !!! trajectory classes is significant,
 R4
    Category > 1   *********  *********      0.850      0.395 !!! despite extreme parameter values.
 R5
    Category > 1       0.000      0.000      1.089      0.276 !!! This is due to large standard errors.
Latent Class 1 Compared to Latent Class 3
 R1
    Category > 1       0.000      0.000      1.052      0.293
 R2
    Category > 1      20.291     18.855      1.076      0.282
 R3
    Category > 1   *********  *********      1.062      0.288
 R4
    Category > 1   *********  *********      1.066      0.286
 R5
    Category > 1       0.000      0.000      0.996      0.319
Latent Class 2 Compared to Latent Class 3
 R1
    Category > 1    6883.415  14441.683      0.477      0.634
 R2
    Category > 1     394.248    369.915      1.066      0.287
 R3
    Category > 1      57.557     55.145      1.044      0.297
 R4
    Category > 1      21.419     17.964      1.192      0.233
 R5
    Category > 1      20.317     11.514      1.765      0.078

QUALITY OF NUMERICAL RESULTS
     Condition Number for the Information Matrix              0.315E-04
       (ratio of smallest to largest eigenvalue)

TECHNICAL 11 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200
     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                           -340.608
          2 Times the Loglikelihood Difference                9.514
          Difference in the Number of Parameters                  4
```

```
          Mean                                              -0.948
          Standard Deviation                                 3.562
          P-Value                                           0.0016
     LO-MENDELL-RUBIN ADJUSTED LRT TEST
          Value                                              9.174
          P-Value                                           0.0019

TECHNICAL 14 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                 1000
        Number of final stage optimizations                    200
     Random Starts Specification for the k-1 Class Model for Generated Data
        Number of initial stage random starts                    0
        Number of final stage optimizations for the
           initial stage random starts                           0
     Random Starts Specification for the k Class Model for Generated Data
        Number of initial stage random starts                   40
        Number of final stage optimizations                      8
     Number of bootstrap draws requested                    Varies

     PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                          -340.608
          2 Times the Loglikelihood Difference               9.514
          Difference in the Number of Parameters                 4
          Approximate P-Value                               0.0505
          Successful Bootstrap Draws                            99

     WARNING:  1 OUT OF 100 BOOTSTRAP DRAWS DID NOT CONVERGE.
     INCREASE THE NUMBER OF RANDOM STARTS USING THE LRTSTARTS OPTION.
     !!! This value for the BLRT matches to 2 decimal places
     !!! that reported in the article. However, there is a minor warning
     !!! regarding convergence.
```


```
MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA  90066
Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com
```

Appendix B: Monte Carlo Power Simulation Study – Derivation Dataset

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
10/03/2015   1:53 PM

!!! Population parameters are set to those produced by Swartout's syntax file (Appendix A)
INPUT INSTRUCTIONS
  TITLE: Derivation Data;
  Monte Carlo simulation study to estimate power
  MONTECARLO:
      NAMES ARE R1 R2 R3 R4 R5;
      NOBSERVATIONS ARE 850;
      NREPS = 240;
      SEED = 53487;
      GENERATE = R1-R5 (1);
      CATEGORICAL ARE R1-R5;
      GENCLASSES = c(3);
      CLASSES =  c(3);
      PATMISS = R1(0) R2(0) R3(1) R4(0) R5(0)| !1
                R1(0) R2(0) R3(0) R4(0) R5(0)| !2
                R1(0) R2(0) R3(0) R4(1) R5(0)| !3
                R1(0) R2(1) R3(1) R4(1) R5(1)| !7
                R1(0) R2(0) R3(1) R4(1) R5(1)| !8
                R1(0) R2(0) R3(0) R4(1) R5(1)| !9
                R1(0) R2(0) R3(0) R4(0) R5(1)| !10
                R1(0) R2(0) R3(1) R4(0) R5(1); !12
      PATPROBS = .06|.07|.04|.25|.28|.18|.10|.02;
  MODEL POPULATION:
      %OVERALL%
      i s q| R1@0 R2@1 R3@2 R4@3 R5@4;
      [R1$1-R5$1@3.251];
      [c#1@-3.948 c#2@-4.189];
      %c#1%
      [i@-59.984 s@79.859 q@-17.574]; !Increasing class
      %c#2%
      [i@8.837 s@-4.194 q@0.624]; !Decreasing class
      %c#3%
      [i@0 s@-.866 q@.156]; !Low or Time-limited class
  ANALYSIS: TYPE = MIXTURE; ALGORITHM = INTEGRATION;
      INTEGRATION = MONTECARLO;
  MODEL:
      %OVERALL%
      i s q| R1@0 R2@1 R3@2 R4@3 R5@4;
        [R1$1-R5$1*3.251];
        [c#1*-3.948 c#2*-4.189];
        %c#1%
        [i*-59.984 s*79.859 q*-17.574]; !Increasing class
        %c#2%
        [i*8.837 s*-4.194 q*0.624]; !Decreasing class
        %c#3%
        [i@0 s*-.866 q*.156]; !Low or Time-limited class
  OUTPUT: TECH9;

INPUT READING TERMINATED NORMALLY

SUMMARY OF ANALYSIS
Number of groups                                                 1
Number of observations                                         850
Number of replications
    Requested                                                  240
    Completed                                                  222

SUMMARY OF DATA FOR THE FIRST REPLICATION
     Number of missing data patterns             8
     Number of y missing data patterns           0
     Number of u missing data patterns           8

SUMMARY OF MISSING DATA PATTERNS FOR THE FIRST REPLICATION
     MISSING DATA PATTERNS FOR U (x = not missing)
           1  2  3  4  5  6  7  8
 R1        x  x  x  x  x  x  x  x
 R2        x  x     x  x  x  x  x
 R3        x        x     x     x
 R4                 x  x     x  x
 R5                    x  x     x

     MISSING DATA PATTERN FREQUENCIES FOR U
    Pattern   Frequency     Pattern   Frequency     Pattern   Frequency
          1         162           4          84           7          15
          2         235           5          55           8          53
          3         221           6          25
```

```
COVARIANCE COVERAGE OF DATA FOR THE FIRST REPLICATION
Minimum covariance coverage value   0.100
     PROPORTION OF DATA PRESENT FOR U
           Covariance Coverage
              R1            R2            R3            R4            R5
              ________      ________      ________      ________      ________
 R1             1.000
 R2             0.740         0.740
 R3             0.381         0.381         0.381
 R4             0.244         0.244         0.161         0.244
 R5             0.156         0.156         0.092         0.127         0.156
     WARNING:  THE COVARIANCE COVERAGE FALLS BELOW THE SPECIFIED LIMIT.

MODEL FIT INFORMATION
!!! The output has been truncated due to space considerations
Number of Free Parameters                        31
Loglikelihood
    H0 Value
        Mean                                 -329.920
        Std Dev                                31.074
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.991         -402.208       -403.358
Information Criteria
    Akaike (AIC)
        Mean                                  721.841
        Std Dev                                62.148
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.986          577.266        570.968
    Bayesian (BIC)
        Mean                                  868.943
        Std Dev                                62.148
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.986          724.368        718.070
    Sample-Size Adjusted BIC (n* = (n + 2) / 24)
        Mean                                  770.496
        Std Dev                                62.148
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.986          625.921        619.623
Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes
    Pearson Chi-Square
        Mean                                    4.693
        Std Dev                                 6.541
        Degrees of freedom                          0
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.968            0.000          0.000
    Likelihood Ratio Chi-Square
        Mean                                    2.002
        Std Dev                                 1.913
        Degrees of freedom                          0
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       1.000            0.000          0.091
Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator Model
    Pearson Chi-Square for MCAR
        Mean                                   38.377
        Std Dev                                11.778
        Degrees of freedom                         63
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.383           39.855         17.882
    Likelihood Ratio Chi-Square
        Mean                                   38.912
        Std Dev                                 8.272
        Degrees of freedom                         63
        Number of successful computations      222
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.419           39.855         21.863

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

```
    Latent
   Classes
       1        13.62842          0.01603
       2        16.45039          0.01935

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES
    Latent
   Classes
       1        13.62371          0.01603
       2        16.45575          0.01936
       3       819.92054          0.96461

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP
Class Counts and Proportions
    Latent
   Classes
       1               7          0.00793 !!! Increasing class
       2              14          0.01658 !!! Decreasing class
       3             829          0.97550 !!! Low or Time-limited class

CLASSIFICATION QUALITY
     Entropy                        0.914

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)
           1        2        3
    1   0.815    0.003    0.120
    2   0.001    0.929    0.070
    3   0.009    0.004    0.986

Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
           1        2        3
    1   0.472    0.000    0.528
    2   0.000    0.685    0.315
    3   0.002    0.001    0.998

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
              1        2        3
    1     -0.111  -13.176    0.000
    2    -12.660    0.777    0.000
    3     -6.470   -7.403    0.000

MODEL RESULTS
!!! Power for detecting each parameter using a p.05 criterion is shown in the last column.
                              ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average   Std. Dev.   Average             Cover Coeff
Latent Class 1
 I        |
  R1                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
 S        |
  R1                  0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  2.000     2.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  3.000     3.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                  4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000
 Q        |
  R1                  0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  9.000     9.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                 16.000    16.0000     0.0000     0.0000     0.0000 1.000 0.000
 S        WITH
  I                   0.000  -196.6457  2627.9788   173.0228 ********** 0.887 0.113
 Q        WITH
  I                   0.000    46.9564   633.7585    34.4270 ********** 0.910 0.090
  S                   0.000  -178.2532  1181.9498   125.3568 ********** 0.896 0.104
 Means
  I                 -59.984  -175.9486   338.4942    16.2501 ********** 0.104 0.982 !!! Power is very high due
  S                  79.859   225.3914   419.4714     5.7502 ********** 0.009 1.000 !!! to the extreme values
  Q                 -17.574   -49.8058    92.9000     0.5616  9630.4307 0.009 1.000 !!! at which they were
fixed
 Thresholds
  R1$1                3.251     3.2510     0.0000     0.0000     0.0000 0.000 1.000
  R2$1                3.251     3.2751     0.3468     0.0000     0.1203 0.000 1.000
  R3$1                3.251     3.0869     0.6744     0.0000     0.4798 0.000 1.000
  R4$1                3.251     3.2510     0.0000     0.0000     0.0000 0.000 1.000
```

```
  R5$1              3.251     3.2510    0.0000     0.0000     0.0000 0.000 1.000
 Variances
  I                 0.050   158.2527  2118.8035   187.5377 ********** 0.860 0.131
  S                 0.050   579.3691  4355.5454   361.0495 ********** 0.874 0.126
  Q                 0.050    65.9007   402.0473    56.4961 ********** 0.865 0.122
Latent Class 2
 I        |
  R1                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R2                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R3                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R4                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R5                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
 S        |
  R1                0.000     0.0000    0.0000     0.0000     0.0000 1.000 0.000
  R2                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R3                2.000     2.0000    0.0000     0.0000     0.0000 1.000 0.000
  R4                3.000     3.0000    0.0000     0.0000     0.0000 1.000 0.000
  R5                4.000     4.0000    0.0000     0.0000     0.0000 1.000 0.000
 Q        |
  R1                0.000     0.0000    0.0000     0.0000     0.0000 1.000 0.000
  R2                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R3                4.000     4.0000    0.0000     0.0000     0.0000 1.000 0.000
  R4                9.000     9.0000    0.0000     0.0000     0.0000 1.000 0.000
  R5               16.000    16.0000    0.0000     0.0000     0.0000 1.000 0.000
 S        WITH
  I                 0.000  -196.6457  2627.9788   173.0228 ********** 0.887 0.113
 Q        WITH
  I                 0.000    46.9564   633.7585    34.4270 ********** 0.910 0.090
  S                 0.000  -178.2532  1181.9498   125.3568 ********** 0.896 0.104
 Means
  I                 8.837   119.8264   480.2127    39.0346 ********** 0.216 0.788 !!! Power is acceptable
  S                -4.194   -69.4158   318.0961    29.0807 ********** 0.212 0.793 !!! or nearly so (at .80)
  Q                 0.624     6.7289    59.7454     5.7096  3590.7053 0.189 0.811 !!! for Decreasing class
 Thresholds                                                                       !!! but the standard error
  R1$1              3.251    37.4233   163.2133     1.7525 27686.3398 0.032 0.968 !!! is improbably high.
  R2$1              3.251    13.6560    71.0404     1.5397  5132.2710 0.045 0.959
  R3$1              3.251    11.6834    77.4279     0.6307  6039.1836 0.005 0.995
  R4$1              3.251     1.5224     7.3182     0.0000    56.3025 0.000 1.000
  R5$1              3.251     3.2581     0.1062     0.0000     0.0113 0.000 1.000
 Variances
  I                 0.050   158.2527  2118.8035   187.5377 ********** 0.860 0.131
  S                 0.050   579.3691  4355.5454   361.0495 ********** 0.874 0.126
  Q                 0.050    65.9007   402.0473    56.4961 ********** 0.865 0.122
Latent Class 3
 I        |
  R1                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R2                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R3                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R4                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R5                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
 S        |
  R1                0.000     0.0000    0.0000     0.0000     0.0000 1.000 0.000
  R2                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R3                2.000     2.0000    0.0000     0.0000     0.0000 1.000 0.000
  R4                3.000     3.0000    0.0000     0.0000     0.0000 1.000 0.000
  R5                4.000     4.0000    0.0000     0.0000     0.0000 1.000 0.000
 Q        |
  R1                0.000     0.0000    0.0000     0.0000     0.0000 1.000 0.000
  R2                1.000     1.0000    0.0000     0.0000     0.0000 1.000 0.000
  R3                4.000     4.0000    0.0000     0.0000     0.0000 1.000 0.000
  R4                9.000     9.0000    0.0000     0.0000     0.0000 1.000 0.000
  R5               16.000    16.0000    0.0000     0.0000     0.0000 1.000 0.000
 S        WITH
  I                 0.000  -196.6457  2627.9788   173.0228 ********** 0.887 0.113
 Q        WITH
  I                 0.000    46.9564   633.7585    34.4270 ********** 0.910 0.090
  S                 0.000  -178.2532  1181.9498   125.3568 ********** 0.896 0.104
 Means
  I                 0.000     0.0000    0.0000     0.0000     0.0000 1.000 0.000 !!! Power is very low
  S                -0.866    -1.4150   86.5120    10.4488  7450.9067 0.572 0.423 !!! for the Low or Time-
  Q                 0.156    -3.0820   23.2668     4.6426   549.3922 0.581 0.410 !!! limited class,
 Thresholds                                                                        !!! which supports the
  R1$1              3.251     8.1811   23.9131    23.2126   593.5674 0.635 0.806 !!! authors’ interpretation
  R2$1              3.251     5.3473   52.8338     9.3362  2783.2275 0.514 0.730 !!! of the class.
  R3$1              3.251     6.3748   68.5307     5.0176  4685.0649 0.374 0.824
  R4$1              3.251     1.0598   17.3282     7.2235   303.7164 0.077 0.923
  R5$1              3.251     3.2458    0.1002    16.5990     0.0100 0.054 0.950
 Variances
  I                 0.050   158.2527  2118.8035   187.5377 ********** 0.860 0.131
  S                 0.050   579.3691  4355.5454   361.0495 ********** 0.874 0.126
  Q                 0.050    65.9007   402.0473    56.4961 ********** 0.865 0.122

Categorical Latent Variables
 Means
```

```
  C#1                -3.948    -5.7702    14.3892     2.1405    209.4383 0.883 0.923 !!! Power is good for
  C#2                -4.189    -4.0664     0.5670     0.4883      0.3351 0.712 0.973 !!! the class separation
                                                                                    !!! but this is likely due
QUALITY OF NUMERICAL RESULTS                                                        !!! to the improbably high
     Average Condition Number for the Information Matrix      0.121E-04             !!! values at which the
       (ratio of smallest to largest eigenvalue)                                    !!! growth parameters were
                                                                                    !!! fixed.
TECHNICAL 9 OUTPUT
!!! Serious warning messages were produced for every iteration. The model is underidentified.
  Error messages for each replication (if any)
    REPLICATION 1:
     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
     THE FOLLOWING PARAMETERS WERE FIXED:
     Parameter 2, %C#1%: [ S ]
     Parameter 11, %C#2%: [ S ]
     Parameter 12, %C#2%: [ Q ]
     Parameter 15, %C#1%: [ R1$1 ]
     Parameter 16, %C#1%: [ R2$1 ]
     Parameter 17, %C#1%: [ R3$1 ]
     Parameter 18, %C#1%: [ R4$1 ]
     Parameter 19, %C#1%: [ R5$1 ]
     Parameter 20, %C#2%: [ R1$1 ]
     Parameter 21, %C#2%: [ R2$1 ]
     Parameter 22, %C#2%: [ R3$1 ]
     Parameter 23, %C#2%: [ R4$1 ]
     Parameter 24, %C#2%: [ R5$1 ]
     Parameter 28, %C#3%: [ R4$1 ]
     Parameter 29, %C#3%: [ R5$1 ]
     Parameter 1, %C#1%: [ I ]
     Parameter 3, %C#1%: [ Q ]

    REPLICATION 2:
     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO AN ILL-CONDITIONED
     FISHER INFORMATION MATRIX.  CHANGE YOUR MODEL AND/OR STARTING VALUES.
     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO A NON-POSITIVE
     DEFINITE FISHER INFORMATION MATRIX.  THIS MAY BE DUE TO THE STARTING VALUES
     BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE CONDITION
     NUMBER IS      -0.506D-09.

    REPLICATION 2:
     THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES COULD NOT BE
     COMPUTED.  THIS IS OFTEN DUE TO THE STARTING VALUES BUT MAY ALSO BE
     AN INDICATION OF MODEL NONIDENTIFICATION.  CHANGE YOUR MODEL AND/OR
     STARTING VALUES.  PROBLEM INVOLVING THE FOLLOWING PARAMETER:
     Parameter 3, %C#1%: [ Q ]

    REPLICATION 3:
     THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE
     TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE
     FIRST-ORDER DERIVATIVE PRODUCT MATRIX.  THIS MAY BE DUE TO THE STARTING
     VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE
     CONDITION NUMBER IS       0.310D-11.  PROBLEM INVOLVING THE FOLLOWING PARAMETER:
     Parameter 12, %C#2%: [ Q ]

    REPLICATION 3:
     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
     THE FOLLOWING PARAMETERS WERE FIXED:
     Parameter 3, %C#1%: [ Q ]
     Parameter 9, %C#1%: Q (equality/label)
     Parameter 13, %C#3%: [ S ]
     Parameter 14, %C#3%: [ Q ]
     Parameter 15, %C#1%: [ R1$1 ]
     Parameter 16, %C#1%: [ R2$1 ]
     Parameter 17, %C#1%: [ R3$1 ]
     Parameter 18, %C#1%: [ R4$1 ]
     Parameter 19, %C#1%: [ R5$1 ]
     Parameter 20, %C#2%: [ R1$1 ]
     Parameter 21, %C#2%: [ R2$1 ]
     Parameter 22, %C#2%: [ R3$1 ]
     Parameter 23, %C#2%: [ R4$1 ]
     Parameter 24, %C#2%: [ R5$1 ]
     Parameter 26, %C#3%: [ R2$1 ]
     Parameter 27, %C#3%: [ R3$1 ]
     Parameter 28, %C#3%: [ R4$1 ]
     Parameter 29, %C#3%: [ R5$1 ]
     Parameter 31, [ C#2 ]
     Parameter 2, %C#1%: [ S ]
```

```
REPLICATION 4:
  THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO AN ILL-CONDITIONED
  FISHER INFORMATION MATRIX.  CHANGE YOUR MODEL AND/OR STARTING VALUES.
  THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO A NON-POSITIVE
  DEFINITE FISHER INFORMATION MATRIX.  THIS MAY BE DUE TO THE STARTING VALUES
  BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE CONDITION
  NUMBER IS      -0.597D-17.

 REPLICATION 4:
  THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES COULD NOT BE
  COMPUTED.  THIS IS OFTEN DUE TO THE STARTING VALUES BUT MAY ALSO BE
  AN INDICATION OF MODEL NONIDENTIFICATION.  CHANGE YOUR MODEL AND/OR
  STARTING VALUES.  PROBLEM INVOLVING THE FOLLOWING PARAMETER:
  Parameter 2, %C#1%: [ S ]

 REPLICATION 5:
  ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
  INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
  MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
  DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
  THE FOLLOWING PARAMETERS WERE FIXED:
  Parameter 3, %C#1%: [ Q ]
  Parameter 11, %C#2%: [ S ]
  Parameter 12, %C#2%: [ Q ]
  Parameter 13, %C#3%: [ S ]
  Parameter 15, %C#1%: [ R1$1 ]
  Parameter 16, %C#1%: [ R2$1 ]
  Parameter 17, %C#1%: [ R3$1 ]
  Parameter 18, %C#1%: [ R4$1 ]
  Parameter 19, %C#1%: [ R5$1 ]
  Parameter 20, %C#2%: [ R1$1 ]
  Parameter 21, %C#2%: [ R2$1 ]
  Parameter 22, %C#2%: [ R3$1 ]
  Parameter 23, %C#2%: [ R4$1 ]
  Parameter 24, %C#2%: [ R5$1 ]
  Parameter 27, %C#3%: [ R3$1 ]
  Parameter 29, %C#3%: [ R5$1 ]
  Parameter 2, %C#1%: [ S ]
      .
      .
      .
 REPLICATION 239:
  ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
  INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
  MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
  DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
  THE FOLLOWING PARAMETERS WERE FIXED:
  Parameter 2, %C#1%: [ S ]
  Parameter 12, %C#2%: [ Q ]
  Parameter 13, %C#3%: [ S ]
  Parameter 14, %C#3%: [ Q ]
  Parameter 15, %C#1%: [ R1$1 ]
  Parameter 16, %C#1%: [ R2$1 ]
  Parameter 17, %C#1%: [ R3$1 ]
  Parameter 18, %C#1%: [ R4$1 ]
  Parameter 19, %C#1%: [ R5$1 ]
  Parameter 20, %C#2%: [ R1$1 ]
  Parameter 21, %C#2%: [ R2$1 ]
  Parameter 22, %C#2%: [ R3$1 ]
  Parameter 23, %C#2%: [ R4$1 ]
  Parameter 24, %C#2%: [ R5$1 ]
  Parameter 27, %C#3%: [ R3$1 ]
  Parameter 28, %C#3%: [ R4$1 ]
  Parameter 29, %C#3%: [ R5$1 ]
  Parameter 11, %C#2%: [ S ]
  Parameter 3, %C#1%: [ Q ]

 REPLICATION 240:
  ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
  INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
  MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
  DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
  THE FOLLOWING PARAMETERS WERE FIXED:
  Parameter 3, %C#1%: [ Q ]
  Parameter 8, %C#1%: Q WITH S (equality/label)
  Parameter 15, %C#1%: [ R1$1 ]
  Parameter 16, %C#1%: [ R2$1 ]
  Parameter 17, %C#1%: [ R3$1 ]
  Parameter 18, %C#1%: [ R4$1 ]
  Parameter 19, %C#1%: [ R5$1 ]
  Parameter 20, %C#2%: [ R1$1 ]
  Parameter 21, %C#2%: [ R2$1 ]
  Parameter 23, %C#2%: [ R4$1 ]
  Parameter 24, %C#2%: [ R5$1 ]
```

```
     Parameter 26, %C#3%: [ R2$1 ]
     Parameter 27, %C#3%: [ R3$1 ]
     Parameter 28, %C#3%: [ R4$1 ]
     Parameter 29, %C#3%: [ R5$1 ]
     Parameter 22, %C#2%: [ R3$1 ]
     Parameter 2, %C#1%: [ S ]

     Beginning Time:  13:53:57
        Ending Time:  16:30:52
       Elapsed Time:  02:36:55
```

Appendix C: Annotated Output Omitting R5 – Derivation Dataset

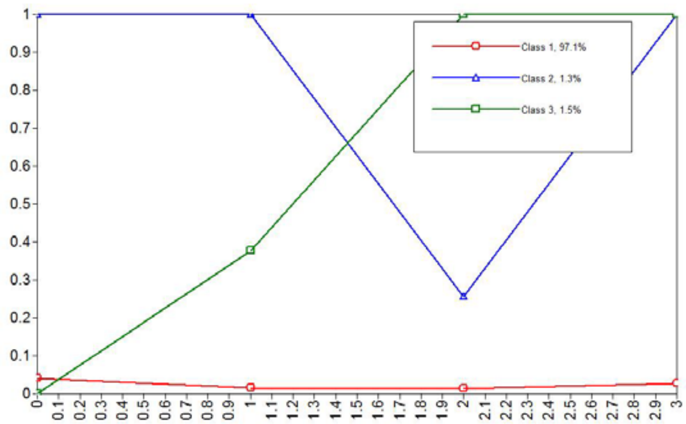
```
Mplus VERSION 7.31
MUTHEN & MUTHEN
09/20/2015   6:27 PM
INPUT INSTRUCTIONS
  TITLE: Final runs of R Traj analysis - Derivation Data – 4-timepoint model;
  DATA: FILE IS RTraj.dat;
  VARIABLE: NAMES ARE id R1 R2 R3 R4 R5 dataset use;
  useobs is (dataset eq 3);
  IDVARIABLE = id;
  USEVAR = R1-R4;
  categorical ARE R1-R4;
  MISSING ARE all(9999999);
  CLASSES =  c(3);
  ANALYSIS: TYPE = MIXTURE;
  Estimator = MLR;
  !for bootstrapped confidence intervals;
  !Estimator = ML;
  !BOOTSTRAP = 15000;
  Processors = 7;
  starts= 1000 200;
  MODEL:
  %OVERALL%
  i s q| R1@0 R2@1 R3@2 R4@3;
  OUTPUT: tech11 tech14 CINTERVAL(BCBOOTSTRAP) ;
  PLOT:
  type = plot3;
  series = R1-R4 (s);

*** WARNING
  Data set contains cases with missing on all variables.
  These cases were not included in the analysis.
  Number of cases with missing on all variables:  1
   2 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS

Final runs of R Traj analysis - Derivation Data;
SUMMARY OF ANALYSIS
Number of observations                                         850

SUMMARY OF DATA
     Number of missing data patterns           10
     Number of y missing data patterns          0
     Number of u missing data patterns         10
COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100
     PROPORTION OF DATA PRESENT FOR U
           Covariance Coverage
              R1            R2            R3            R4
              ________      ________      ________      ________
 R1            0.996
 R2            0.745         0.748
 R3            0.404         0.392         0.405
 R4            0.265         0.261         0.182         0.267

UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES
    R1
      Category 1    0.948          803.000
      Category 2    0.052           44.000
    R2
      Category 1    0.969          616.000
      Category 2    0.031           20.000
    R3
      Category 1    0.968          333.000
      Category 2    0.032           11.000
    R4
      Category 1    0.956          217.000
      Category 2    0.044           10.000

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES
Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
            -325.812  383979           603
            -325.812  345070           114
            -325.812  377584           630
            -325.812  590834           785
            -325.812  383902           673
            -325.812  502157           799
            .
            .
            .
            -329.451  722748           346
            -329.451  97158            205
            -329.451  370466           41
```

```
           -329.589  695155           150
           -329.589  937225           394
           -329.589  348637           749
           -329.589  736574           414
           -329.589  898745           466
           -329.888  94573            983
           -329.952  792993           859
           -329.955  783165           170
THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED.  RERUN WITH AT LEAST TWICE THE
RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.

     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
     THE FOLLOWING PARAMETERS WERE FIXED:
     Parameter 6, %C#2%: [ S ]  !!! These represent the most critical growth
     Parameter 9, %C#3%: [ Q ]  !!! parameters for the two smallest classes.
     Parameter 7, %C#2%: [ Q ]
     Parameter 8, %C#3%: [ S ]

THE MODEL ESTIMATION TERMINATED NORMALLY
MODEL FIT INFORMATION
Number of Free Parameters                        11
Loglikelihood
          H0 Value                         -325.812
          H0 Scaling Correction Factor       0.6359
            for MLR
Information Criteria
          Akaike (AIC)                      673.624
          Bayesian (BIC)                    725.821
          Sample-Size Adjusted BIC          690.889
            (n* = (n + 2) / 24)
Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes
          Pearson Chi-Square
          Value                               1.620
          Degrees of Freedom                      4
          P-Value                            0.8051
          Likelihood Ratio Chi-Square
          Value                               1.062
          Degrees of Freedom                      4
          P-Value                            0.9002
Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator Model
          Pearson Chi-Square
          Value                              57.319
          Degrees of Freedom                     39
          P-Value                            0.0294  !!! In this model, the MCAR/MAR assumption
                                                     !!! is shown to be untenable.
          Likelihood Ratio Chi-Square
          Value                              24.667
          Degrees of Freedom                     39
          P-Value                            0.9642

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL
    Latent
   Classes
       1        825.58442          0.97128
       2         11.31625          0.01331
       3         13.09933          0.01541

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES
    Latent
   Classes
       1        825.58440          0.97128
       2         11.31625          0.01331
       3         13.09935          0.01541

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP
Class Counts and Proportions
    Latent
   Classes
       1              836          0.98353  !!! Low or Time-limited class
       2                8          0.00941  !!! Decreasing class
       3                6          0.00706  !!! Increasing class

CLASSIFICATION QUALITY
     Entropy                         0.951

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)
```

```
           1         2         3
    1   0.986     0.004     0.010
    2   0.035     0.965     0.000
    3   0.231     0.000     0.769

Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
           1         2         3
    1   0.998     0.000     0.002
    2   0.318     0.682     0.000  !!! Strong misclassification with Low or Time-limited class
    3   0.648     0.000     0.352

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
              1         2         3
    1      6.386    -1.607     0.000
    2     12.669    13.433     0.000
    3      0.610   -12.772     0.000


MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Latent Class 1 !!! Low or Time-limited
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
 Means
    I                 32.084      1.067     30.081      0.000  !!! All three growth parameters
    S                 -1.435      0.536     -2.679      0.007  !!! of Low or Time-limited
    Q                  0.433      0.188      2.300      0.021  !!! class are significant.
 Thresholds
    R1$1              35.271      1.041     33.888      0.000
    R2$1              35.271      1.041     33.888      0.000
    R3$1              35.271      1.041     33.888      0.000
    R4$1              35.271      1.041     33.888      0.000

Latent Class 2  !!! Decreasing class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
 Means
    I                139.143      1.565     88.933      0.000  !!! Improbably high values
    S               -104.579      0.000    999.000    999.000  !!! for the intercept and
    Q                 26.055      0.000    999.000    999.000  !!! slope trajectory parameters;
                                                               !!! the slope parameter
                                                               !!! was fixed.
 Thresholds
    R1$1              35.271      1.041     33.888      0.000  !!! Improbably high threshold values
    R2$1              35.271      1.041     33.888      0.000
    R3$1              35.271      1.041     33.888      0.000
    R4$1              35.271      1.041     33.888      0.000

Latent Class 3 !!! Increasing class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
 S        |
```

```
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
 Means
    I                  0.000      0.000    999.000    999.000  !!! All growth parameters are
    S                 41.726      0.000    999.000    999.000  !!! fixed. The intercept of the
    Q                 -6.960      0.000    999.000    999.000  !!! last class is fixed to zero
                                                               !!! by default in Mplus.
 Thresholds
    R1$1              35.271      1.041     33.888      0.000
    R2$1              35.271      1.041     33.888      0.000
    R3$1              35.271      1.041     33.888      0.000
    R4$1              35.271      1.041     33.888      0.000

Categorical Latent Variables
 Means
    C#1                4.144      0.533      7.778      0.000
    C#2               -0.146      0.632     -0.231      0.817 !!! The Increasing and Decreasing classes
                                                              !!! are not statistically indistinguishable.

RESULTS IN PROBABILITY SCALE
Latent Class 1
 R1
    Category 1         0.960      0.007    132.437      0.000
    Category 2         0.040      0.007      5.472      0.000
 R2
    Category 1         0.985      0.005    212.409      0.000
    Category 2         0.015      0.005      3.222      0.001
 R3
    Category 1         0.987      0.004    222.365      0.000
    Category 2         0.013      0.004      2.943      0.003
 R4
    Category 1         0.973      0.012     80.833      0.000
    Category 2         0.027      0.012      2.218      0.027
Latent Class 2
 R1
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
 R2
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
 R3
    Category 1         0.743      0.223      3.337      0.001
    Category 2         0.257      0.223      1.152      0.249
 R4
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
Latent Class 3
 R1
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 R2
    Category 1         0.624      0.244      2.552      0.011
    Category 2         0.376      0.244      1.541      0.123
 R3
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
 R4
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000

LATENT CLASS ODDS RATIO RESULTS  !!! None of these comparisons are significant.
Latent Class 1 Compared to Latent Class 2
 R1
    Category > 1       0.000      0.000      0.846      0.397
 R2
    Category > 1       0.000      0.000      0.829      0.407
 R3
    Category > 1       0.038      0.047      0.823      0.411
 R4
    Category > 1       0.000      0.000      0.796      0.426

Latent Class 1 Compared to Latent Class 3
 R1
    Category > 1   *********  *********      0.938      0.348
 R2
    Category > 1       0.025      0.028      0.912      0.362
 R3
```

```
    Category > 1       0.000      0.000      0.940      0.347
 R4
    Category > 1       0.000      0.000      0.977      0.329

Latent Class 2 Compared to Latent Class 3
 R1
    Category > 1   *********  *********      0.639      0.523
 R2
    Category > 1   *********  *********      0.639      0.523
 R3
    Category > 1       0.000      0.000      0.639      0.523
 R4
    Category > 1       0.072      0.112      0.639      0.523

QUALITY OF NUMERICAL RESULTS
     Condition Number for the Information Matrix              0.467E-03
       (ratio of smallest to largest eigenvalue)

TECHNICAL 11 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200
     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                         -330.962
          2 Times the Loglikelihood Difference             10.300
          Difference in the Number of Parameters                4
          Mean                                             -0.240
          Standard Deviation                                3.107
          P-Value                                          0.0015
     LO-MENDELL-RUBIN ADJUSTED LRT TEST
          Value                                             9.932
          P-Value                                          0.0018

TECHNICAL 14 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200
     Random Starts Specification for the k-1 Class Model for Generated Data
        Number of initial stage random starts                   0
        Number of final stage optimizations for the
           initial stage random starts                          0
     Random Starts Specification for the k Class Model for Generated Data
        Number of initial stage random starts                  40
        Number of final stage optimizations                     8
     Number of bootstrap draws requested                   Varies

     PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                         -330.962
          2 Times the Loglikelihood Difference             10.300
          Difference in the Number of Parameters                4
          Approximate P-Value                              0.0000
          Successful Bootstrap Draws                           49

     WARNING:  3 OUT OF 52 BOOTSTRAP DRAWS DID NOT CONVERGE.
     INCREASE THE NUMBER OF RANDOM STARTS USING THE LRTSTARTS OPTION.
!!! This indicates a problem in the bootstrapping, perhaps due to empty cells.

     Beginning Time:  18:27:37
        Ending Time:  18:30:04
       Elapsed Time:  00:02:27

MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA  90066
Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com
Copyright (c) 1998-2015 Muthen & Muthen
```


Class 1, 97.1%
Class 2, 1.3%
Class 3, 1.5%
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3

Appendix D: Annotated Output Generated by Swartout's Syntax Files – Validation Dataset

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
09/20/2015   5:51 PM
INPUT INSTRUCTIONS
  TITLE: Final runs of R Traj analysis - Validation Data;
  DATA: FILE IS RTraj.dat;
  VARIABLE: NAMES ARE id R1 R2 R3 R4 R5 dataset use;
  useobs is (dataset eq 4);
  IDVARIABLE = id;
  USEVAR = R1-R5;
  categorical ARE R1-R5;
  MISSING ARE all(9999999);
  CLASSES =  c(3);
  ANALYSIS: TYPE = MIXTURE;
  Estimator = MLR;
  !for bootstrapped confidence intervals;
  !Estimator = ML;
  !BOOTSTRAP = 15000;
  Processors = 7;
  starts= 1000 200;
  MODEL:
  %OVERALL%
  i s q| R1@0 R2@1 R3@2 R4@3 R5@4;
  OUTPUT: tech11 tech14 CINTERVAL(BCBOOTSTRAP);
  PLOT:
  type = plot3;
  series = R1-R5 (s);
  !!! Most of the same issues noted above are also evident in the validation model.
  !!! It is somewhat better identified, likely due to less missing data.
  !!! There are a few more cases assigned to the smallest classes.


Final runs of R Traj analysis - Validation Data;
SUMMARY OF ANALYSIS
Number of observations                                          795

SUMMARY OF DATA
     Number of missing data patterns            10
     Number of y missing data patterns           0
     Number of u missing data patterns          10

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100

     PROPORTION OF DATA PRESENT FOR U
           Covariance Coverage
              R1            R2            R3            R4            R5
              ________      ________      ________      ________      ________
 R1             0.995
 R2             0.995         0.997
 R3             0.806         0.809         0.811
 R4             0.747         0.750         0.682         0.752
 R5             0.707         0.709         0.647         0.634         0.712

UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES
    R1
      Category 1    0.949          751.000
      Category 2    0.051           40.000
    R2
      Category 1    0.956          758.000
      Category 2    0.044           35.000
    R3
      Category 1    0.936          604.000
      Category 2    0.064           41.000
    R4
      Category 1    0.957          572.000
      Category 2    0.043           26.000
    R5
      Category 1    0.958          542.000
      Category 2    0.042           24.000

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES
Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
            -571.029  667250           318
            -571.029  922596           456
            -571.029  348637           749
            -571.029  629320           222
            -571.029  193042           316
            -571.029  497522           502
            -571.029  314084           81
            -571.029  437181           135
```

```
            -571.029  794236           127
            .
            .
            .
            -572.811  354624           448
            -572.811  576220           115
            -572.811  980970           894
            -572.811  496710           386
            -572.811  494149           815
            -572.811  79945            395
            -572.811  551340           766
            -572.811  391368           802

THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED.  RERUN WITH AT LEAST TWICE THE
RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.

THE MODEL ESTIMATION TERMINATED NORMALLY
MODEL FIT INFORMATION
Number of Free Parameters                       11
Loglikelihood
          H0 Value                        -571.029
          H0 Scaling Correction Factor      1.0270
            for MLR
Information Criteria
          Akaike (AIC)                    1164.058
          Bayesian (BIC)                  1215.520
          Sample-Size Adjusted BIC        1180.589
            (n* = (n + 2) / 24)
            !!! Like with the derivation model, this adjusted BIC is not exactly what
            !!! is reported in the article.

Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes**
          Pearson Chi-Square
          Value                             12.807
          Degrees of Freedom                    19
          P-Value                           0.8483
          Likelihood Ratio Chi-Square
          Value                             16.132
          Degrees of Freedom                    19
          P-Value                           0.6485

** Of the 132 cells in the latent class indicator table, 1
   were deleted in the calculation of chi-square due to extreme values.

Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator Model
          Pearson Chi-Square
          Value                             54.349
          Degrees of Freedom                    91
          P-Value                           0.9992
          Likelihood Ratio Chi-Square
          Value                             47.120
          Degrees of Freedom                    91
          P-Value                           1.0000

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL
    Latent
   Classes
       1         44.60518          0.05611
       2        728.90684          0.91686
       3         21.48799          0.02703
       !!! This is the value that is reported in the article.
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES
    Latent
   Classes
       1         44.60521          0.05611
       2        728.90681          0.91686
       3         21.48798          0.02703

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP
Class Counts and Proportions
    Latent
   Classes
       1               39          0.04906  !!! Decreasing class
       2              737          0.92704  !!! Low or Time-limited class
       3               19          0.02390  !!! Increasing class

CLASSIFICATION QUALITY
     Entropy                         0.918

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
```

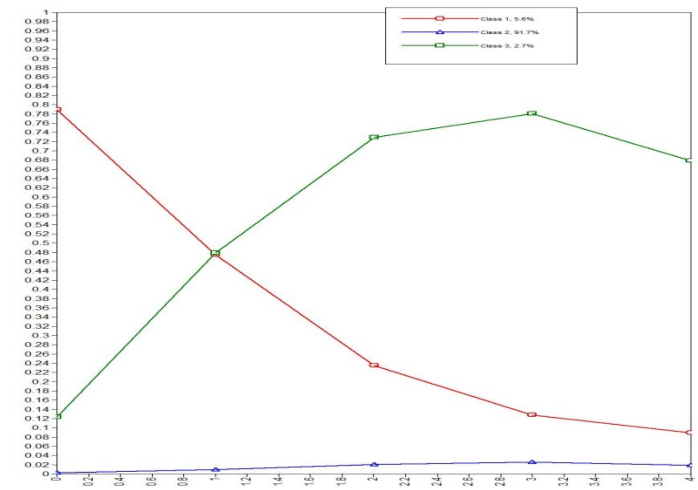
```
by Latent Class (Column)
           1        2        3
    1   0.907    0.050    0.043 !!! This matches the range reported in the article
    2   0.010    0.984    0.006 !!! The classification matrix below shows a moderate
    3   0.104    0.076    0.820 !!! amount of misclassification of both Class 1 and 3
                                !!! (Decreasing & Increasing, respectively) with
                                !!! Class 2 (Low or time-limited).
Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
           1        2        3
    1   0.793    0.163    0.044
    2   0.003    0.995    0.002
    3   0.079    0.196    0.725

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
              1        2        3
    1      2.887    1.306    0.000
    2      0.305    6.222    0.000
    3     -2.223   -1.309    0.000

MODEL RESULTS

                                                  Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Latent Class 1  !!! Decreasing class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                  3.280      1.762      1.862      0.063  !!! None are significant
    S                 -1.594      0.855     -1.865      0.062
    Q                  0.171      0.186      0.919      0.358
 Thresholds
    R1$1               1.959      1.489      1.316      0.188
    R2$1               1.959      1.489      1.316      0.188
    R3$1               1.959      1.489      1.316      0.188
    R4$1               1.959      1.489      1.316      0.188
    R5$1               1.959      1.489      1.316      0.188

Latent Class 2  !!! Low or Time-limited class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                 -3.985      1.964     -2.029      0.042 !!! Only the intercept growth
    S                  1.607      0.874      1.839      0.066 !!! parameter (pre-college rape)
    Q                 -0.278      0.171     -1.630      0.103 !!! is significant.

 Thresholds
    R1$1               1.959      1.489      1.316      0.188
    R2$1               1.959      1.489      1.316      0.188
    R3$1               1.959      1.489      1.316      0.188
    R4$1               1.959      1.489      1.316      0.188
    R5$1               1.959      1.489      1.316      0.188
```

```
Latent Class 3  !!! Increasing class
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                  0.000      0.000    999.000    999.000
    S                  2.276      1.155      1.970      0.049  !!! The slope growth parameter
    Q                 -0.400      0.219     -1.826      0.068  !!! is significant.

 Thresholds
    R1$1               1.959      1.489      1.316      0.188
    R2$1               1.959      1.489      1.316      0.188
    R3$1               1.959      1.489      1.316      0.188
    R4$1               1.959      1.489      1.316      0.188
    R5$1               1.959      1.489      1.316      0.188

Categorical Latent Variables
 Means
    C#1                0.730      0.504      1.449      0.147 !!! The Decreasing class is not
    C#2                3.524      0.355      9.924      0.000 !!! distinguishable from the
                                                              !!! Increasing class.
RESULTS IN PROBABILITY SCALE

Latent Class 1
 R1
    Category 1         0.211      0.142      1.482      0.138
    Category 2         0.789      0.142      5.555      0.000
 R2
    Category 1         0.525      0.091      5.764      0.000
    Category 2         0.475      0.091      5.205      0.000
 R3
    Category 1         0.766      0.071     10.710      0.000
    Category 2         0.234      0.071      3.279      0.001
 R4
    Category 1         0.873      0.051     17.023      0.000
    Category 2         0.127      0.051      2.487      0.013
 R5
    Category 1         0.911      0.059     15.360      0.000
    Category 2         0.089      0.059      1.507      0.132

Latent Class 2
 R1
    Category 1         0.997      0.003    364.258      0.000
    Category 2         0.003      0.003      0.955      0.340
 R2
    Category 1         0.990      0.004    233.969      0.000
    Category 2         0.010      0.004      2.316      0.021
 R3
    Category 1         0.979      0.006    173.167      0.000
    Category 2         0.021      0.006      3.708      0.000
 R4
    Category 1         0.974      0.006    161.489      0.000
    Category 2         0.026      0.006      4.288      0.000
 R5
    Category 1         0.981      0.007    148.772      0.000
    Category 2         0.019      0.007      2.807      0.005

Latent Class 3
 R1
    Category 1         0.876      0.161      5.436      0.000
    Category 2         0.124      0.161      0.766      0.443
 R2
    Category 1         0.521      0.174      2.996      0.003
    Category 2         0.479      0.174      2.756      0.006
 R3
    Category 1         0.270      0.100      2.698      0.007
    Category 2         0.730      0.100      7.281      0.000
 R4
```

```
    Category 1         0.219      0.097      2.272      0.023
    Category 2         0.781      0.097      8.083      0.000
 R5
    Category 1         0.322      0.169      1.905      0.057
    Category 2         0.678      0.169      4.016      0.000

LATENT CLASS ODDS RATIO RESULTS
Latent Class 1 Compared to Latent Class 2
 R1
    Category > 1    1429.879   1541.095      0.928      0.353
 R2
    Category > 1      91.238     39.638      2.302      0.021
 R3
    Category > 1      14.298      7.519      1.902      0.057
 R4
    Category > 1       5.503      3.114      1.767      0.077
 R5
    Category > 1       5.201      4.296      1.211      0.226

Latent Class 1 Compared to Latent Class 3
 R1
    Category > 1      26.582     46.829      0.568      0.570
 R2
    Category > 1       0.981      0.882      1.113      0.266
 R3
    Category > 1       0.113      0.075      1.509      0.131
 R4
    Category > 1       0.041      0.027      1.509      0.131
 R5
    Category > 1       0.047      0.051      0.907      0.365

Latent Class 2 Compared to Latent Class 3
 R1
    Category > 1       0.019      0.037      0.509      0.611
 R2
    Category > 1       0.011      0.009      1.196      0.232
 R3
    Category > 1       0.008      0.004      1.874      0.061
 R4
    Category > 1       0.007      0.004      1.777      0.076
 R5
    Category > 1       0.009      0.007      1.195      0.232

QUALITY OF NUMERICAL RESULTS
     Condition Number for the Information Matrix              0.434E-04
       (ratio of smallest to largest eigenvalue)

TECHNICAL 11 OUTPUT

     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200

     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                           -581.368
          2 Times the Loglikelihood Difference               20.679
          Difference in the Number of Parameters                  4
          Mean                                                0.064
          Standard Deviation                                 11.075
          P-Value                                            0.0240
     LO-MENDELL-RUBIN ADJUSTED LRT TEST
          Value                                              19.932
          P-Value                                            0.0266

TECHNICAL 14 OUTPUT

     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200
     Random Starts Specification for the k-1 Class Model for Generated Data
        Number of initial stage random starts                   0
        Number of final stage optimizations for the
           initial stage random starts                          0
     Random Starts Specification for the k Class Model for Generated Data
        Number of initial stage random starts                  40
        Number of final stage optimizations                     8
     Number of bootstrap draws requested                   Varies

     PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                           -581.368
          2 Times the Loglikelihood Difference               20.679
          Difference in the Number of Parameters                  4
          Approximate P-Value                                0.0000
          Successful Bootstrap Draws                             49
```

```
     Beginning Time:  17:51:00
        Ending Time:  17:53:21
       Elapsed Time:  00:02:21
```


!!! The graph produced does not exactly match the one shown in the JAMA article.



Appendix E: Monte Carlo Power Simulation Study – Validation Dataset

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
10/04/2015  12:24 PM

INPUT INSTRUCTIONS
  TITLE: Validation Data;
  Monte Carlo simulation study to estimate power of class separation
  MONTECARLO:
      NAMES ARE R1 R2 R3 R4 R5;
      NOBSERVATIONS ARE 795;
      NREPS = 250;
      SEED = 53487;
      GENERATE = R1-R5 (1);
      CATEGORICAL ARE R1-R5;
      GENCLASSES = c(3);
      CLASSES =  c(3);
      PATMISS = R1(0) R2(0) R3(1) R4(0) R5(0)|
                R1(0) R2(0) R3(0) R4(0) R5(0)|
                R1(0) R2(0) R3(0) R4(1) R5(0)|
                R1(0) R2(0) R3(1) R4(1) R5(0)|
                R1(0) R2(0) R3(0) R4(1) R5(1)|
                R1(0) R2(0) R3(0) R4(0) R5(1)|
                R1(0) R2(0) R3(1) R4(1) R5(1)|
                R1(0) R2(0) R3(1) R4(0) R5(1);
      PATPROBS = .05|.58|.07|.02|.07|.09|.10|.02;
  MODEL POPULATION:
      %OVERALL%
      i s q| R1@0 R2@1 R3@2 R4@3 R5@4;
      [R1$1-R5$1@1.959];
      [c#1@0.730 c#2@3.524];
      %c#1%
      [i@3.28 s@-1.594 q@0.171]; !Decreasing class
      %c#2%
      [i@-3.985 s@1.607 q@-0.278]; !Low or Time-limited class
      %c#3%
      [i@0 s@2.276 q@-0.400]; !Increasing class
  ANALYSIS: TYPE = MIXTURE MISSING; ALGORITHM = INTEGRATION;
      INTEGRATION = MONTECARLO;
  MODEL:
      %OVERALL%
      i s q| R1@0 R2@1 R3@2 R4@3 R5@4;
      [R1$1-R5$1*1.959];
      [c#1*0.730 c#2*3.524];
      %c#1%
      [i*3.28 s*-1.594 q*0.171]; !Decreasing class
      %c#2%
      [i*-3.985 s*1.607 q*-0.278]; !Low or Time-limited class
      %c#3%
      [i*0 s*2.276 q*-0.400]; !Increasing class
  OUTPUT: TECH9;

SUMMARY OF ANALYSIS
Number of observations                                          795
Number of replications
    Requested                                                   250
    Completed                                                   182

SUMMARY OF DATA FOR THE FIRST REPLICATION
     Number of missing data patterns            8
     Number of y missing data patterns          0
     Number of u missing data patterns          8

SUMMARY OF MISSING DATA PATTERNS FOR THE FIRST REPLICATION
     MISSING DATA PATTERNS FOR U (x = not missing)

           1  2  3  4  5  6  7  8
 R1        x  x  x  x  x  x  x  x
 R2        x  x  x  x  x  x  x  x
 R3        x  x  x  x
 R4        x     x           x  x
 R5           x  x     x     x

     MISSING DATA PATTERN FREQUENCIES FOR U

    Pattern   Frequency     Pattern   Frequency     Pattern   Frequency
          1          62           4          73           7          44
          2          61           5          18           8          13
          3         445           6          79
COVARIANCE COVERAGE OF DATA FOR THE FIRST REPLICATION
Minimum covariance coverage value   0.100
```

```
     PROPORTION OF DATA PRESENT FOR U
           Covariance Coverage
              R1            R2            R3            R4            R5
              ________      ________      ________      ________      ________
 R1             1.000
 R2             1.000         1.000
 R3             0.806         0.806         0.806
 R4             0.709         0.709         0.638         0.709
 R5             0.714         0.714         0.636         0.615         0.714

MODEL FIT INFORMATION
Number of Free Parameters                        32
Loglikelihood
    H0 Value
        Mean                               -563.164
        Std Dev                              39.677
        Number of successful computations      182
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.989         -655.464       -676.090
Information Criteria
    Akaike (AIC)
        Mean                               1190.329
        Std Dev                              79.353
        Number of successful computations      182
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.984         1005.729        930.769
    Bayesian (BIC)
        Mean                               1340.036
        Std Dev                              79.353
        Number of successful computations      182
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.984         1155.436       1080.476
    Sample-Size Adjusted BIC (n* = (n + 2) / 24)
        Mean                               1238.418
        Std Dev                              79.353
        Number of successful computations      182
             Proportions                   Percentiles
        Expected    Observed         Expected       Observed
           0.990       0.984         1053.819        978.859
           0.010       0.011         1423.017       1400.891

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

    Latent
   Classes
       1        45.70395          0.05749
       2       729.82335          0.91802
       3        19.47270          0.02449

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES

    Latent
   Classes
       1        45.71710          0.05751
       2       729.81113          0.91800
       3        19.47177          0.02449
FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

    Latent
   Classes
       1              39          0.04949
       2             739          0.92986
       3              16          0.02066

CLASSIFICATION QUALITY

     Entropy                         0.947

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)

           1        2        3

    1   0.934    0.042    0.023
    2   0.011    0.983    0.005
    3   0.049    0.064    0.887
```

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

           1        2        3

    1   0.901    0.094    0.006
    2   0.007    0.993    0.000
    3   0.059    0.073    0.868
Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)

              1        2        3

    1      5.058    2.795    0.000
    2      4.257    9.285    0.000
    3     -2.690   -2.477    0.000

MODEL RESULTS
!!! Power for detecting all trajectory means is < .60.
                              ESTIMATES              S. E.     M. S. E.  95%  % Sig
                 Population   Average   Std. Dev.   Average             Cover Coeff
Latent Class 1 !!! Decreasing class

 I        |
  R1                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000

 S        |
  R1                  0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  2.000     2.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  3.000     3.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                  4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000

 Q        |
  R1                  0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  9.000     9.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                 16.000    16.0000     0.0000     0.0000     0.0000 1.000 0.000

 S        WITH
  I                   0.000   -44.2930   148.7933   187.0491 23979.6543 0.984 0.016

 Q        WITH
  I                   0.000     9.9635    35.8091    42.5760  1374.5203 0.978 0.022
  S                   0.000   -12.4334    46.4055    75.7109  2296.2234 0.978 0.022

 Means
  I                   3.280     2.4813    91.8075    85.7060  8382.9355 0.632 0.341
  S                  -1.594     2.1118    70.6067    58.4186  4971.6455 0.643 0.346
  Q                   0.171    -1.5948    14.2609     9.8332   205.3726 0.643 0.363

 Thresholds
  R1$1                1.959   -69.7573   752.1437    51.1774 ********** 0.533 0.451
  R2$1                1.959     2.7796    37.3663    23.7611  1389.2440 0.615 0.385
  R3$1                1.959     3.0922     9.7381     0.0000    95.5929 0.000 1.000
  R4$1                1.959     1.7428     1.0763     0.0000     1.1989 0.000 1.000
  R5$1                1.959     1.9597     0.0225     0.0000     0.0005 0.000 1.000

 Variances
  I                   0.050    40.0762   118.9414   202.1509 15671.4277 0.978 0.011
  S                   0.050    53.1356   191.0991   301.5703 39136.2812 0.978 0.022
  Q                   0.050     3.0190    11.4253    20.7500   138.6351 0.967 0.027

Latent Class 2 !!! Low or Time-limited class

 I        |
  R1                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000

 S        |
  R1                  0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                  1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                  2.000     2.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                  3.000     3.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                  4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000
```

```
 Q        |
  R1                 0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                 4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                 9.000     9.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                16.000    16.0000     0.0000     0.0000     0.0000 1.000 0.000

 S        WITH
  I                  0.000   -44.2930   148.7933   187.0491 23979.6543 0.984 0.016

 Q        WITH
  I                  0.000     9.9635    35.8091    42.5760  1374.5203 0.978 0.022
  S                  0.000   -12.4334    46.4055    75.7109  2296.2234 0.978 0.022

 Means
  I                 -3.985   -21.3091    42.4707   103.8899  2093.9746 0.852 0.269
  S                  1.607    14.1867    31.2370    80.5237  1128.6364 0.874 0.176
  Q                 -0.278    -2.6976     5.7408    17.4282    38.6301 0.879 0.154

 Thresholds
  R1$1               1.959     9.2647    27.1275   112.7363   785.2323 0.560 0.456
  R2$1               1.959    -2.5816    17.7694    19.9332   334.6349 0.780 0.187
  R3$1               1.959     1.8766     2.1112     0.0185     4.4393 0.011 0.989
  R4$1               1.959     2.2037     0.8088     0.0216     0.7104 0.005 0.995
  R5$1               1.959     1.9890     0.1459     0.0000     0.0221 0.000 1.000

 Variances
  I                  0.050    40.0762   118.9414   202.1509 15671.4277 0.978 0.011
  S                  0.050    53.1356   191.0991   301.5703 39136.2812 0.978 0.022
  Q                  0.050     3.0190    11.4253    20.7500   138.6351 0.967 0.027

Latent Class 3 !!! Increasing class

 I        |
  R1                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000

 S        |
  R1                 0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                 2.000     2.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                 3.000     3.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                 4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000

 Q        |
  R1                 0.000     0.0000     0.0000     0.0000     0.0000 1.000 0.000
  R2                 1.000     1.0000     0.0000     0.0000     0.0000 1.000 0.000
  R3                 4.000     4.0000     0.0000     0.0000     0.0000 1.000 0.000
  R4                 9.000     9.0000     0.0000     0.0000     0.0000 1.000 0.000
  R5                16.000    16.0000     0.0000     0.0000     0.0000 1.000 0.000

 S        WITH
  I                  0.000   -44.2930   148.7933   187.0491 23979.6543 0.984 0.016

 Q        WITH
  I                  0.000     9.9635    35.8091    42.5760  1374.5203 0.978 0.022
  S                  0.000   -12.4334    46.4055    75.7109  2296.2234 0.978 0.022

 Means
  I                  0.000   -25.0151   253.2460    63.1442 64406.9141 0.445 0.555
  S                  2.276    37.7593   286.6295    58.3323 82964.1172 0.456 0.533
  Q                 -0.400    -7.2640    56.2796    11.5974  3197.1086 0.462 0.527

 Thresholds
  R1$1               1.959     5.2460   247.5200    44.0326 60940.3125 0.231 0.769
  R2$1               1.959     3.4352    47.5194     9.1098  2247.8674 0.313 0.698
  R3$1               1.959    11.1282   103.4605     0.1197 10729.3389 0.005 0.995
  R4$1               1.959     2.9479     4.1145     3.1962    17.8141 0.005 0.995
  R5$1               1.959     1.9605     0.0532     1.4509     0.0028 0.011 0.989

 Variances
  I                  0.050    40.0762   118.9414   202.1509 15671.4277 0.978 0.011
  S                  0.050    53.1356   191.0991   301.5703 39136.2812 0.978 0.022
  Q                  0.050     3.0190    11.4253    20.7500   138.6351 0.967 0.027

Categorical Latent Variables
!!! Power to distinguish between Increasing and Decreasing trajectory classes is .44.
 Means
  C#1                0.730     0.9047     0.7805     2.0229     0.6364 0.747 0.440
  C#2                3.524     3.7592     0.5223     1.1459     0.3266 0.775 0.901
```

```
QUALITY OF NUMERICAL RESULTS

     Average Condition Number for the Information Matrix      0.297E-06
       (ratio of smallest to largest eigenvalue)

TECHNICAL 9 OUTPUT

  Error messages for each replication (if any)
!!! Problems with estimation for each of the 250 replications
     REPLICATION 1:
     THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE
     TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE
     FIRST-ORDER DERIVATIVE PRODUCT MATRIX.  THIS MAY BE DUE TO THE STARTING
     VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE
     CONDITION NUMBER IS       0.108D-10.  PROBLEM INVOLVING THE FOLLOWING PARAMETER:
     Parameter 16, %C#1%: [ R1$1 ]

     REPLICATION 1:
     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
     THE FOLLOWING PARAMETERS WERE FIXED:
     Parameter 18, %C#1%: [ R3$1 ]
     Parameter 19, %C#1%: [ R4$1 ]
     Parameter 20, %C#1%: [ R5$1 ]
     Parameter 21, %C#2%: [ R1$1 ]
     Parameter 22, %C#2%: [ R2$1 ]
     Parameter 23, %C#2%: [ R3$1 ]
     Parameter 24, %C#2%: [ R4$1 ]
     Parameter 25, %C#2%: [ R5$1 ]
     Parameter 27, %C#3%: [ R2$1 ]
     Parameter 28, %C#3%: [ R3$1 ]
     Parameter 29, %C#3%: [ R4$1 ]
     Parameter 30, %C#3%: [ R5$1 ]
     THE DEGREES OF FREEDOM FOR THIS MODEL ARE NEGATIVE.  THE MODEL IS NOT
     IDENTIFIED OR TOO MANY CELLS WERE DELETED.  A CHI-SQUARE TEST IS NOT
     AVAILABLE.
     REPLICATION 2:
     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO AN ILL-CONDITIONED
     FISHER INFORMATION MATRIX.  CHANGE YOUR MODEL AND/OR STARTING VALUES.

     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO A NON-POSITIVE
     DEFINITE FISHER INFORMATION MATRIX.  THIS MAY BE DUE TO THE STARTING VALUES
     BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE CONDITION
     NUMBER IS      -0.363D-16.

     REPLICATION 2:
     THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES COULD NOT BE
     COMPUTED.  THIS IS OFTEN DUE TO THE STARTING VALUES BUT MAY ALSO BE
     AN INDICATION OF MODEL NONIDENTIFICATION.  CHANGE YOUR MODEL AND/OR
     STARTING VALUES.  PROBLEM INVOLVING THE FOLLOWING PARAMETER:
     Parameter 18, %C#1%: [ R3$1 ]

     REPLICATION 3:
     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
     THE FOLLOWING PARAMETERS WERE FIXED:
     Parameter 18, %C#1%: [ R3$1 ]
     Parameter 19, %C#1%: [ R4$1 ]
     Parameter 20, %C#1%: [ R5$1 ]
     Parameter 21, %C#2%: [ R1$1 ]
     Parameter 23, %C#2%: [ R3$1 ]
     Parameter 24, %C#2%: [ R4$1 ]
     Parameter 25, %C#2%: [ R5$1 ]
     Parameter 26, %C#3%: [ R1$1 ]
     Parameter 27, %C#3%: [ R2$1 ]
     Parameter 28, %C#3%: [ R3$1 ]
     Parameter 29, %C#3%: [ R4$1 ]
     Parameter 30, %C#3%: [ R5$1 ]
     THE DEGREES OF FREEDOM FOR THIS MODEL ARE NEGATIVE.  THE MODEL IS NOT
     IDENTIFIED OR TOO MANY CELLS WERE DELETED.  A CHI-SQUARE TEST IS NOT
     AVAILABLE.
         .
         .
         .
     REPLICATION 248:
     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.
     THE FOLLOWING PARAMETERS WERE FIXED:
```

```
     Parameter 16, %C#1%: [ R1$1 ]
     Parameter 17, %C#1%: [ R2$1 ]
     Parameter 19, %C#1%: [ R4$1 ]
     Parameter 20, %C#1%: [ R5$1 ]
     Parameter 23, %C#2%: [ R3$1 ]
     Parameter 24, %C#2%: [ R4$1 ]
     Parameter 25, %C#2%: [ R5$1 ]
     Parameter 26, %C#3%: [ R1$1 ]
     Parameter 27, %C#3%: [ R2$1 ]
     Parameter 28, %C#3%: [ R3$1 ]
     Parameter 29, %C#3%: [ R4$1 ]
     Parameter 30, %C#3%: [ R5$1 ]
     Parameter 18, %C#1%: [ R3$1 ]
     THE DEGREES OF FREEDOM FOR THIS MODEL ARE NEGATIVE.  THE MODEL IS NOT
     IDENTIFIED OR TOO MANY CELLS WERE DELETED.  A CHI-SQUARE TEST IS NOT
     AVAILABLE.
     REPLICATION 249:
     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO AN ILL-CONDITIONED
     FISHER INFORMATION MATRIX.  CHANGE YOUR MODEL AND/OR STARTING VALUES.

     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO A NON-POSITIVE
     DEFINITE FISHER INFORMATION MATRIX.  THIS MAY BE DUE TO THE STARTING VALUES
     BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE CONDITION
     NUMBER IS      -0.370D-16.

     REPLICATION 249:
     THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES COULD NOT BE
     COMPUTED.  THIS IS OFTEN DUE TO THE STARTING VALUES BUT MAY ALSO BE
     AN INDICATION OF MODEL NONIDENTIFICATION.  CHANGE YOUR MODEL AND/OR
     STARTING VALUES.  PROBLEM INVOLVING THE FOLLOWING PARAMETER:
     Parameter 18, %C#1%: [ R3$1 ]

     REPLICATION 250:
     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO AN ILL-CONDITIONED
     FISHER INFORMATION MATRIX.  CHANGE YOUR MODEL AND/OR STARTING VALUES.

     THE MODEL ESTIMATION DID NOT TERMINATE NORMALLY DUE TO A NON-POSITIVE
     DEFINITE FISHER INFORMATION MATRIX.  THIS MAY BE DUE TO THE STARTING VALUES
     BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION.  THE CONDITION
     NUMBER IS      -0.437D-10.

     REPLICATION 250:
     THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES COULD NOT BE
     COMPUTED.  THIS IS OFTEN DUE TO THE STARTING VALUES BUT MAY ALSO BE
     AN INDICATION OF MODEL NONIDENTIFICATION.  CHANGE YOUR MODEL AND/OR
     STARTING VALUES.  PROBLEM INVOLVING THE FOLLOWING PARAMETER:
     Parameter 18, %C#1%: [ R3$1 ]

DIAGRAM INFORMATION

  Mplus diagrams are currently not available for Mixture analysis.
  No diagram output was produced.
     Beginning Time:  12:24:38
        Ending Time:  17:46:45
       Elapsed Time:  05:22:07

MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA  90066

Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com
```

Appendix F: Rapists Only (Both Datasets Combined) – Latent Trajectory Analysis

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
09/05/2015   4:02 PM
INPUT INSTRUCTIONS
  TITLE: Derivation and validation data combined (rapists only);

  DATA: FILE IS RTraj_rapists only.dat;
  VARIABLE:
      NAMES ARE id R1 R2 R3 R4 R5 dataset use;
      IDVARIABLE = id;
      USEVAR = R1-R5;
      CATEGORICAL = R1-R5;
      MISSING ARE R1-R5(9);
      CLASSES =  c(3);
  ANALYSIS: TYPE = MIXTURE; ESTIMATOR = MLR; PROCESSORS = 7;
    STARTS = 1000 200;
  MODEL:
    %OVERALL%
    i s q| R1@0 R2@1 R3@2 R4@3 R5@4;
  OUTPUT: TECH11 TECH14 CINTERVAL(BCBOOTSTRAP);
  PLOT: TYPE = PLOT3; SERIES = R1-R5 (s);


Derivation and validation data combined (rapists only);
SUMMARY OF ANALYSIS
Number of observations                                        127 !!! Combined the samples
                                                                  !!! due to the small sample sizes.


SUMMARY OF DATA
     Number of missing data patterns             9
     Number of y missing data patterns           0
     Number of u missing data patterns           9

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100

     PROPORTION OF DATA PRESENT FOR U

           Covariance Coverage
              R1            R2            R3            R4            R5
              ________      ________      ________      ________      ________
 R1             0.992
 R2             0.969         0.976
 R3             0.787         0.772         0.795
 R4             0.724         0.732         0.630         0.732
 R5             0.614         0.622         0.535         0.583         0.622

UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES
    R1
      Category 1    0.730           92.000 !!! Among rapists, the proportions admitting to
      Category 2    0.270           34.000 !!! rape at each timepoint is higher, making
    R2                                     !!! computation easier.
      Category 1    0.556           69.000
      Category 2    0.444           55.000
    R3
      Category 1    0.485           49.000
      Category 2    0.515           52.000
    R4
      Category 1    0.613           57.000
      Category 2    0.387           36.000
    R5
      Category 1    0.684           54.000
      Category 2    0.316           25.000

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES
Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
            -321.994  694303           282
            -321.994  859432           770
            -321.994  507154           387
            -321.994  291149           536
            -321.994  302046           863
            -321.994  22075            659
            -321.994  440841           118
            -321.994  751153           110
            -321.994  596257           405
            -321.994  982520           737
            -321.994  544009           842
            -321.994  177175           851
            -321.994  30098            209
```

```
            .
            .
            .
            -325.838  937885           426
            -325.838  788796           145
            -325.838  642386           662
            -325.838  535804           111
            -325.838  535303           923
            -325.838  562716           300
            -326.420  793035           187
            -327.448  599729           658
            -327.448  741888           138
            -328.054  609185           181
            -328.054  903369           134
            -328.054  674171           195

THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED.  RERUN WITH AT LEAST TWICE THE
RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.

     ONE OR MORE PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE
     INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE
     MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT
     DISTRIBUTION OF THE CATEGORICAL VARIABLES IN THE MODEL.

     THE FOLLOWING PARAMETERS WERE FIXED:
     Parameter 9, %C#3%: [ Q ]
     Parameter 8, %C#3%: [ S ]
  !!! As with previous latent variable growth mixture models, the model is misspecified.

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION
Number of Free Parameters                        11
Loglikelihood
          H0 Value                          -321.994
          H0 Scaling Correction Factor        0.7209
            for MLR
Information Criteria
          Akaike (AIC)                       665.989
          Bayesian (BIC)                     697.275
          Sample-Size Adjusted BIC           662.488
            (n* = (n + 2) / 24)
Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes
          Pearson Chi-Square
          Value                               95.779
          Degrees of Freedom                      20
          P-Value                             0.0000
          Likelihood Ratio Chi-Square
          Value                               60.584
          Degrees of Freedom                      20
          P-Value                             0.0000
Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator Model
          Pearson Chi-Square
          Value                              103.589
          Degrees of Freedom                      80
          P-Value                             0.0393
   !!! Evidence that missing data is related to rape.

          Likelihood Ratio Chi-Square
          Value                               99.915
          Degrees of Freedom                      80
          P-Value                             0.0653

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL
    Latent
   Classes
       1        77.56287          0.61073
       2        27.52290          0.21672
       3        21.91423          0.17255

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES
    Latent
   Classes
       1        77.56287          0.61073
       2        27.52290          0.21672
       3        21.91423          0.17255

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP
Class Counts and Proportions
    Latent
```

```
   Classes
       1               78          0.61417
       2               25          0.19685
       3               24          0.18898
```
!!! Smallest class is far larger than 5%, although the number of members is still smaller than optimal.

```
CLASSIFICATION QUALITY
     Entropy                         0.705
```
!!! Entropy is low.

```
Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)
           1        2        3
    1   0.933    0.041    0.025
    2   0.072    0.814    0.114
    3   0.124    0.164    0.712
```
!!! Evidence of some misclassification

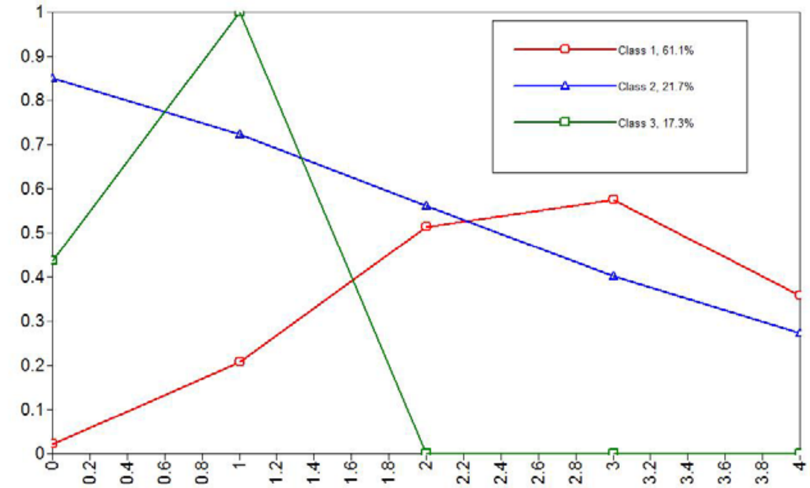
```
Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
           1        2        3
    1   0.938    0.023    0.038
    2   0.117    0.740    0.143
    3   0.091    0.130    0.779

Logits for the Classification Probabilities for the Most Likely Latent Class Membership (Column)
by Latent Class (Row)
              1        2        3
    1      3.195   -0.511    0.000
    2     -0.200    1.643    0.000
    3     -2.152   -1.791    0.000

MODEL RESULTS
                                                       Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Latent Class 1
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                 -3.609      1.039     -3.475      0.001  !!! Only Class 1 (largest class)
    S                  3.101      0.757      4.098      0.000  !!! can be described by a smooth
    Q                 -0.570      0.147     -3.875      0.000  !!! quadratic trajectory.
 Thresholds
    R1$1               0.258      0.621      0.416      0.677
    R2$1               0.258      0.621      0.416      0.677
    R3$1               0.258      0.621      0.416      0.677
    R4$1               0.258      0.621      0.416      0.677
    R5$1               0.258      0.621      0.416      0.677

Latent Class 2
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                  1.997      1.661      1.202      0.229 !!! None of these growth parameters
```

```
    S                 -0.813      1.310     -0.621      0.535 !!! are significant.
    Q                  0.033      0.262      0.126      0.899
 Thresholds
    R1$1               0.258      0.621      0.416      0.677
    R2$1               0.258      0.621      0.416      0.677
    R3$1               0.258      0.621      0.416      0.677
    R4$1               0.258      0.621      0.416      0.677
    R5$1               0.258      0.621      0.416      0.677

Latent Class 3
 I        |
    R1                 1.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 1.000      0.000    999.000    999.000
    R4                 1.000      0.000    999.000    999.000
    R5                 1.000      0.000    999.000    999.000
 S        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 2.000      0.000    999.000    999.000
    R4                 3.000      0.000    999.000    999.000
    R5                 4.000      0.000    999.000    999.000
 Q        |
    R1                 0.000      0.000    999.000    999.000
    R2                 1.000      0.000    999.000    999.000
    R3                 4.000      0.000    999.000    999.000
    R4                 9.000      0.000    999.000    999.000
    R5                16.000      0.000    999.000    999.000
 Means
    I                  0.000      0.000    999.000    999.000 !!! All these growth parameters
    S                 66.807      0.000    999.000    999.000 !!! were fixed to avoid singularity.
    Q                -39.825      0.000    999.000    999.000
 Thresholds
    R1$1               0.258      0.621      0.416      0.677
    R2$1               0.258      0.621      0.416      0.677
    R3$1               0.258      0.621      0.416      0.677
    R4$1               0.258      0.621      0.416      0.677
    R5$1               0.258      0.621      0.416      0.677

Categorical Latent Variables
 Means
    C#1                1.264      0.347      3.640      0.000
    C#2                0.228      0.463      0.492      0.623

RESULTS IN PROBABILITY SCALE
Latent Class 1
 R1
    Category 1         0.980      0.018     54.291      0.000
    Category 2         0.020      0.018      1.136      0.256
 R2
    Category 1         0.792      0.057     13.897      0.000
    Category 2         0.208      0.057      3.653      0.000
 R3
    Category 1         0.486      0.048     10.229      0.000
    Category 2         0.514      0.048     10.798      0.000
 R4
    Category 1         0.425      0.040     10.591      0.000
    Category 2         0.575      0.040     14.353      0.000
 R5
    Category 1         0.643      0.065      9.895      0.000
    Category 2         0.357      0.065      5.503      0.000
Latent Class 2
 R1
    Category 1         0.149      0.175      0.856      0.392
    Category 2         0.851      0.175      4.874      0.000
 R2
    Category 1         0.277      0.085      3.259      0.001
    Category 2         0.723      0.085      8.505      0.000
 R3
    Category 1         0.439      0.092      4.759      0.000
    Category 2         0.561      0.092      6.086      0.000
 R4
    Category 1         0.599      0.102      5.878      0.000
    Category 2         0.401      0.102      3.936      0.000
 R5
    Category 1         0.727      0.120      6.073      0.000
    Category 2         0.273      0.120      2.276      0.023
Latent Class 3
 R1
    Category 1         0.564      0.153      3.694      0.000
    Category 2         0.436      0.153      2.853      0.004
 R2
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
```

```
 R3
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 R4
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 R5
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000

LATENT CLASS ODDS RATIO RESULTS
Latent Class 1 Compared to Latent Class 2
 R1
    Category > 1       0.004      0.005      0.741      0.459
 R2
    Category > 1       0.101      0.044      2.273      0.023
 R3
    Category > 1       0.825      0.371      2.225      0.026
 R4
    Category > 1       2.024      0.927      2.183      0.029
 R5
    Category > 1       1.484      1.030      1.440      0.150
Latent Class 1 Compared to Latent Class 3
 R1
    Category > 1       0.027      0.028      0.963      0.336
 R2
    Category > 1       0.000      0.000      1.455      0.146
 R3
    Category > 1   *********  *********      1.515      0.130
 R4
    Category > 1   *********  *********      1.515      0.130
 R5
    Category > 1   *********  *********      1.472      0.141
Latent Class 2 Compared to Latent Class 3
 R1
    Category > 1       7.368     12.237      0.602      0.547
 R2
    Category > 1       0.000      0.000      1.222      0.222
 R3
    Category > 1   *********  *********      1.554      0.120
 R4
    Category > 1   *********  *********      1.570      0.116
 R5
    Category > 1   *********  *********      1.186      0.235

QUALITY OF NUMERICAL RESULTS
     Condition Number for the Information Matrix              0.302E-04
       (ratio of smallest to largest eigenvalue)

TECHNICAL 11 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200

     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                         -328.511
          2 Times the Loglikelihood Difference             13.034
          Difference in the Number of Parameters                4
          Mean                                              0.601
          Standard Deviation                                3.565
          P-Value                                          0.0015
     LO-MENDELL-RUBIN ADJUSTED LRT TEST
          Value                                            12.394
          P-Value                                          0.0020

TECHNICAL 14 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200
     Random Starts Specification for the k-1 Class Model for Generated Data
        Number of initial stage random starts                   0
        Number of final stage optimizations for the
           initial stage random starts                          0
     Random Starts Specification for the k Class Model for Generated Data
        Number of initial stage random starts                  40
        Number of final stage optimizations                     8
     Number of bootstrap draws requested                   Varies

     PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES
          H0 Loglikelihood Value                         -328.511
          2 Times the Loglikelihood Difference             13.034
          Difference in the Number of Parameters                4
          Approximate P-Value                              0.0300
          Successful Bootstrap Draws                          100
```

```
     Beginning Time:  16:02:41
        Ending Time:  16:03:26
       Elapsed Time:  00:00:45
```

MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA 90066
Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com



Appendix G: Rapists Only (Both Datasets Combined) – Latent Profile Analysis

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
09/05/2015   4:20 PM

INPUT INSTRUCTIONS
  TITLE: Derivation and validation data - Rapists only, no trajectories;
  DATA: FILE IS RTraj_rapists only.dat;
  VARIABLE:
      NAMES = id R1 R2 R3 R4 R5 dataset use;
      IDVARIABLE = id;
      USEVAR = R1-R5;
      CATEGORICAL = R1-R5;
      MISSING = R1-R5(9);
      CLASSES =  c(5);
  ANALYSIS: TYPE = MIXTURE; ESTIMATOR = MLR; PROCESSORS = 7;
      STARTS= 1000 200;
  MODEL:
      %OVERALL%
 !!! No trajectories are specified.
 !!! Only probabilities of rape at each timepoint are estimated.
  OUTPUT: TECH11 TECH14 CINTERVAL(BCBOOTSTRAP);
  PLOT: TYPE = PLOT3; SERIES = R1(0) R2(1) R3(2) R4(3) R5(4);

Final runs of R Traj analysis - Derivation Data;

SUMMARY OF ANALYSIS
Number of observations                                         127

SUMMARY OF DATA
     Number of missing data patterns             9

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100
     PROPORTION OF DATA PRESENT FOR U
           Covariance Coverage
              R1            R2            R3            R4            R5
              ________      ________      ________      ________      ________
 R1             0.992
 R2             0.969         0.976
 R3             0.787         0.772         0.795
 R4             0.724         0.732         0.630         0.732
 R5             0.614         0.622         0.535         0.583         0.622

UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES
    R1
      Category 1    0.730          92.000
      Category 2    0.270          34.000
    R2
      Category 1    0.556          69.000
      Category 2    0.444          55.000
    R3
      Category 1    0.485          49.000
      Category 2    0.515          52.000
    R4
      Category 1    0.613          57.000
      Category 2    0.387          36.000
    R5
      Category 1    0.684          54.000
      Category 2    0.316          25.000

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES
Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
            -295.207  466971           109
            -295.207  889774           954
            -295.207  830570           369
            -295.207  12477            155
            -295.207  118438           601
            -295.207  574942           558
            -295.207  109357           765
            .
            .
            .
            -300.320  136842           58
            -300.354  494209           904
            -303.174  605565           404
            -303.174  275475           413
            -303.174  944186           541
THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED.  RERUN WITH AT LEAST TWICE THE
RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.
     IN THE OPTIMIZATION, ONE OR MORE LOGIT THRESHOLDS APPROACHED AND WERE SET
```

```
     AT THE EXTREME VALUES.  EXTREME VALUES ARE -15.000 AND 15.000.
     THE FOLLOWING THRESHOLDS WERE SET AT THESE VALUES:
     * THRESHOLD 1 OF CLASS INDICATOR R2 FOR CLASS 1 AT ITERATION 89
     * THRESHOLD 1 OF CLASS INDICATOR R3 FOR CLASS 3 AT ITERATION 89
  !!! The output shows that many more parameters than these were fixed at the extreme values.
  !!! Evidence of computational problems, even when describing probabilities only.

THE MODEL ESTIMATION TERMINATED NORMALLY
MODEL FIT INFORMATION
Number of Free Parameters                       29
Loglikelihood
          H0 Value                        -295.207
          H0 Scaling Correction Factor      1.0005
            for MLR
Information Criteria
          Akaike (AIC)                     648.413
          Bayesian (BIC)                   730.895
          Sample-Size Adjusted BIC         639.184
            (n* = (n + 2) / 24)
Chi-Square Test of Model Fit for the Binary and Ordered Categorical
(Ordinal) Outcomes
          Pearson Chi-Square
          Value                              8.551
          Degrees of Freedom                     2
          P-Value                           0.0139
          Likelihood Ratio Chi-Square
          Value                              7.008
          Degrees of Freedom                     2
          P-Value                           0.0301
Chi-Square Test for MCAR under the Unrestricted Latent Class Indicator Model
          Pearson Chi-Square
          Value                            103.589
          Degrees of Freedom                    80
          P-Value                           0.0393 !!! MCAR (b/c no covariates, also MAR)
          Likelihood Ratio Chi-Square              !!! assumption not tenable
          Value                             99.915
          Degrees of Freedom                    80
          P-Value                           0.0653

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL
    Latent
   Classes
       1         15.47910          0.12188
       2         30.66955          0.24149
       3         36.84191          0.29009
       4         15.49076          0.12197
       5         28.51867          0.22456

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON ESTIMATED POSTERIOR PROBABILITIES
    Latent
   Classes
       1         15.47910          0.12188
       2         30.66955          0.24149
       3         36.84191          0.29009
       4         15.49076          0.12197
       5         28.51867          0.22456

FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP
Class Counts and Proportions
    Latent
   Classes
       1               12          0.09449  !!! The proportions of the classes are somewhat
       2               28          0.22047  !!! more evenly distributed.
       3               42          0.33071
       4               14          0.11024
       5               31          0.24409

CLASSIFICATION QUALITY
     Entropy                              0.875 !!! Somewhat lower than optimal

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row)
by Latent Class (Column)
           1         2         3         4         5
    1   0.933     0.000     0.000     0.007     0.061
    2   0.000     1.000     0.000     0.000     0.000
    3   0.089     0.000     0.877     0.015     0.019
    4   0.000     0.016     0.000     0.984     0.000
    5   0.017     0.079     0.000     0.033     0.871

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

	1	2	3	4	5
1	0.723	0.000	0.242	0.000	0.035
2	0.000	0.913	0.000	0.007	0.080
3	0.000	0.000	1.000	0.000	0.000
4	0.005	0.000	0.040	0.889	0.065
5	0.025	0.000	0.028	0.000	0.947

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

	1	2	3	4	5
1	3.034	-10.457	1.941	-10.457	0.000
2	-11.286	2.438	-11.286	-2.387	0.000
3	0.000	0.000	13.816	0.000	0.000
4	-2.498	-11.089	-0.496	2.609	0.000
5	-3.615	-13.761	-3.532	-13.761	0.000

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
Thresholds				
R1$1	0.099	0.623	0.159	0.873
R2$1	-15.000	0.000	999.000	999.000
R3$1	-15.000	0.000	999.000	999.000
R4$1	-0.544	0.777	-0.701	0.483
R5$1	-15.000	0.000	999.000	999.000
Latent Class 2				
Thresholds				
R1$1	3.187	1.011	3.152	0.002
R2$1	15.000	0.000	999.000	999.000
R3$1	0.309	0.426	0.724	0.469
R4$1	-15.000	0.000	999.000	999.000
R5$1	1.419	0.549	2.585	0.010
Latent Class 3				
Thresholds				
R1$1	-0.010	0.366	-0.027	0.978
R2$1	-15.000	0.000	999.000	999.000
R3$1	15.000	0.000	999.000	999.000
R4$1	1.855	0.646	2.871	0.004
R5$1	2.422	1.176	2.059	0.040
Latent Class 4				
Thresholds				
R1$1	2.585	1.051	2.460	0.014
R2$1	3.038	4.677	0.650	0.516
R3$1	1.585	1.025	1.546	0.122
R4$1	15.000	0.000	999.000	999.000
R5$1	-15.000	0.000	999.000	999.000
Latent Class 5				
Thresholds				
R1$1	1.261	0.486	2.593	0.010
R2$1	2.277	0.974	2.339	0.019
R3$1	-15.000	0.000	999.000	999.000
R4$1	15.000	0.000	999.000	999.000
R5$1	15.000	0.000	999.000	999.000
Categorical Latent Variables				
Means				
C#1	-0.611	0.386	-1.584	0.113
C#2	0.073	0.288	0.252	0.801
C#3	0.256	0.286	0.897	0.370
C#4	-0.610	0.407	-1.500	0.134

!!! None of these are distinguishable
!!! from the final (arbitrarily chosen)
!!! class.

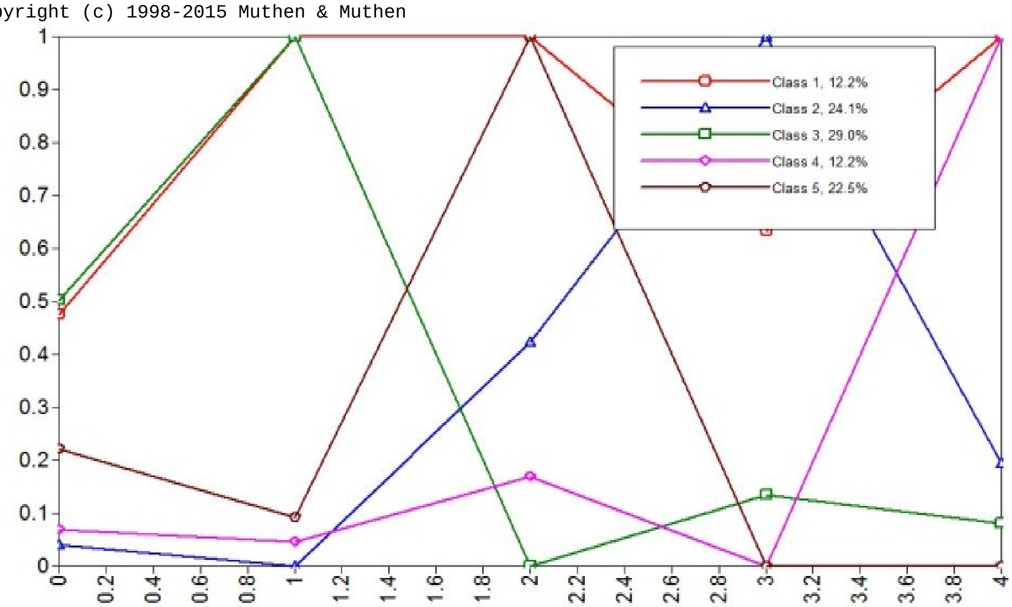
RESULTS IN PROBABILITY SCALE

	Estimate	S.E.	Est./S.E.	P-Value
Latent Class 1				
R1				
Category 1	0.525	0.155	3.379	0.001
Category 2	0.475	0.155	3.060	0.002
R2				
Category 1	0.000	0.000	0.000	1.000
Category 2	1.000	0.000	0.000	1.000
R3				
Category 1	0.000	0.000	0.000	1.000
Category 2	1.000	0.000	0.000	1.000
R4				
Category 1	0.367	0.180	2.035	0.042
Category 2	0.633	0.180	3.507	0.000
R5				
Category 1	0.000	0.000	0.000	1.000
Category 2	1.000	0.000	0.000	1.000
Latent Class 2				
R1				
Category 1	0.960	0.039	24.944	0.000
Category 2	0.040	0.039	1.030	0.303
R2				

```
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 R3
    Category 1         0.577      0.104      5.538      0.000
    Category 2         0.423      0.104      4.067      0.000
 R4
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
 R5
    Category 1         0.805      0.086      9.351      0.000
    Category 2         0.195      0.086      2.262      0.024
Latent Class 3
 R1
    Category 1         0.498      0.091      5.442      0.000
    Category 2         0.502      0.091      5.496      0.000
 R2
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
 R3
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 R4
    Category 1         0.865      0.076     11.442      0.000
    Category 2         0.135      0.076      1.790      0.073
 R5
    Category 1         0.918      0.088     10.428      0.000
    Category 2         0.082      0.088      0.926      0.355
Latent Class 4
 R1
    Category 1         0.930      0.069     13.572      0.000
    Category 2         0.070      0.069      1.023      0.306
 R2
    Category 1         0.954      0.204      4.674      0.000
    Category 2         0.046      0.204      0.224      0.823
 R3
    Category 1         0.830      0.145      5.735      0.000
    Category 2         0.170      0.145      1.176      0.240
 R4
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 R5
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
Latent Class 5
 R1
    Category 1         0.779      0.084      9.312      0.000
    Category 2         0.221      0.084      2.639      0.008
 R2
    Category 1         0.907      0.082     11.041      0.000
    Category 2         0.093      0.082      1.133      0.257
 R3
    Category 1         0.000      0.000      0.000      1.000
    Category 2         1.000      0.000      0.000      1.000
 R4
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000
 R5
    Category 1         1.000      0.000      0.000      1.000
    Category 2         0.000      0.000      0.000      1.000

LATENT CLASS ODDS RATIO RESULTS !!! Most of these comparisons are inestimable.
Latent Class 1 Compared to Latent Class 2
 R1
    Category > 1      21.939     26.036      0.843      0.399
 R2
    Category > 1   *********      0.000    999.000    999.000
 R3
    Category > 1   *********      0.000    999.000    999.000
 R4
    Category > 1       0.000      0.000    999.000    999.000
 R5
    Category > 1   *********      0.000    999.000    999.000
Latent Class 1 Compared to Latent Class 3
 R1
    Category > 1       0.897      0.677      1.324      0.186
 R2
    Category > 1       1.000      0.000    999.000    999.000
 R3
    Category > 1   *********      0.000    999.000    999.000
 R4
    Category > 1      11.018     10.925      1.008      0.313
 R5
    Category > 1   *********      0.000    999.000    999.000
Latent Class 1 Compared to Latent Class 4
```

```
 R1
    Category > 1      12.011     14.648      0.820      0.412
 R2
    Category > 1   *********      0.000    999.000    999.000
 R3
    Category > 1   *********      0.000    999.000    999.000
 R4
    Category > 1   *********      0.000    999.000    999.000
 R5
    Category > 1       1.000      0.000    999.000    999.000
Latent Class 1 Compared to Latent Class 5
 R1
    Category > 1       3.195      2.596      1.231      0.218
 R2
    Category > 1   *********      0.000    999.000    999.000
 R3
    Category > 1       1.000      0.000    999.000    999.000
 R4
    Category > 1   *********      0.000    999.000    999.000
 R5
    Category > 1   *********      0.000    999.000    999.000
Latent Class 2 Compared to Latent Class 3
 R1
    Category > 1       0.041      0.044      0.930      0.352
 R2
    Category > 1       0.000      0.000    999.000    999.000
 R3
    Category > 1   *********      0.000    999.000    999.000
 R4
    Category > 1   *********      0.000    999.000    999.000
 R5
    Category > 1       2.725      3.538      0.770      0.441
Latent Class 2 Compared to Latent Class 4
 R1
    Category > 1       0.547      0.800      0.685      0.494
 R2
    Category > 1       0.000      0.000    999.000    999.000
 R3
    Category > 1       3.582      4.000      0.896      0.370
 R4
    Category > 1   *********      0.000    999.000    999.000
 R5
    Category > 1       0.000      0.000    999.000    999.000
Latent Class 2 Compared to Latent Class 5
 R1
    Category > 1       0.146      0.168      0.865      0.387
 R2
    Category > 1       0.000      0.000    999.000    999.000
 R3
    Category > 1       0.000      0.000    999.000    999.000
 R4
    Category > 1   *********      0.000    999.000    999.000
 R5
    Category > 1   *********      0.000    999.000    999.000
Latent Class 3 Compared to Latent Class 4
 R1
    Category > 1      13.397     14.799      0.905      0.365
 R2
    Category > 1   *********      0.000    999.000    999.000
 R3
    Category > 1       0.000      0.000    999.000    999.000
 R4
    Category > 1   *********      0.000    999.000    999.000
 R5
    Category > 1       0.000      0.000    999.000    999.000
Latent Class 3 Compared to Latent Class 5
 R1
    Category > 1       3.564      2.175      1.639      0.101
 R2
    Category > 1   *********      0.000    999.000    999.000
 R3
    Category > 1       0.000      0.000    999.000    999.000
 R4
    Category > 1   *********      0.000    999.000    999.000
 R5
    Category > 1   *********      0.000    999.000    999.000
Latent Class 4 Compared to Latent Class 5
 R1
    Category > 1       0.266      0.311      0.854      0.393
 R2
    Category > 1       0.467      2.229      0.210      0.834
 R3
    Category > 1       0.000      0.000    999.000    999.000
 R4
```

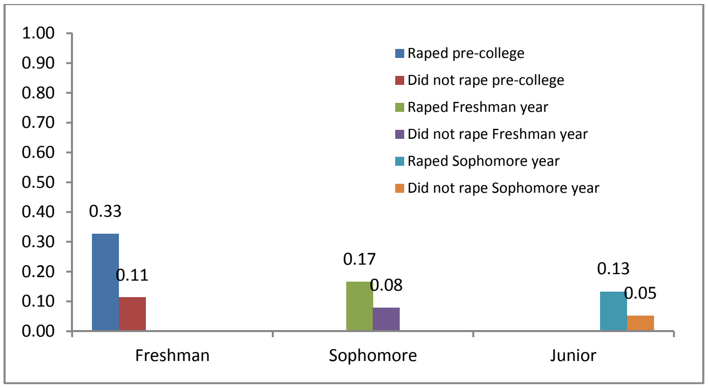
```
    Category > 1       1.000      0.000    999.000    999.000
 R5
    Category > 1   *********      0.000    999.000    999.000

QUALITY OF NUMERICAL RESULTS
     Condition Number for the Information Matrix              0.145E-02
       (ratio of smallest to largest eigenvalue)

TECHNICAL 11 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200
     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 4 (H0) VERSUS 5 CLASSES
          H0 Loglikelihood Value                         -303.346
          2 Times the Loglikelihood Difference             16.279
          Difference in the Number of Parameters                6
          Mean                                             -1.158
          Standard Deviation                                7.158
          P-Value                                          0.0025
     LO-MENDELL-RUBIN ADJUSTED LRT TEST
          Value                                            15.738
          P-Value                                          0.0030

TECHNICAL 14 OUTPUT
     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                1000
        Number of final stage optimizations                   200
     Random Starts Specification for the k-1 Class Model for Generated Data
        Number of initial stage random starts                   0
        Number of final stage optimizations for the
           initial stage random starts                          0
     Random Starts Specification for the k Class Model for Generated Data
        Number of initial stage random starts                  40
        Number of final stage optimizations                     8
     Number of bootstrap draws requested                   Varies

     PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 4 (H0) VERSUS 5 CLASSES
          H0 Loglikelihood Value                         -303.346
          2 Times the Loglikelihood Difference             16.279
          Difference in the Number of Parameters                6
          Approximate P-Value                              0.0000
          Successful Bootstrap Draws                           49

     Beginning Time:  16:20:25
        Ending Time:  16:20:50
       Elapsed Time:  00:00:25
```

Appendix H: Autoregression (No Latent Classes) – Derivation Dataset

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
09/27/2015   6:32 PM

INPUT INSTRUCTIONS
  TITLE: Autoregression - Derivation Data;
  DATA: FILE IS RTraj.dat;
  VARIABLE: NAMES ARE id R1 R2 R3 R4 R5 dataset use;
  useobs is (dataset eq 3);
  IDVARIABLE = id;
  USEVAR = R1-R4;
  categorical ARE R2-R4;
  MISSING ARE all(9999999);

  ANALYSIS: TYPE = GENERAL; PARAMETERIZATION=THETA;

  MODEL:
  R4 ON R3;
  R3 ON R2;
  R2 ON R1;

  MODEL INDIRECT:
  R4 IND R1;
  R4 IND R2;
  R3 IND R1;

*** WARNING
  Data set contains cases with missing on all variables.
  These cases were not included in the analysis.
  Number of cases with missing on all variables:  1
*** WARNING
  Data set contains cases with missing on x-variables.
  These cases were not included in the analysis.
  Number of cases with missing on x-variables:  3
*** WARNING
  Data set contains cases with missing on all variables except
  x-variables.  These cases were not included in the analysis.
  Number of cases with missing on all variables except x-variables:  203
!!! These individuals dropped out of the study after the first survey administration so none of the
!!! college years were represented in their data.

   3 WARNING(S) FOUND IN THE INPUT INSTRUCTIONS

SUMMARY OF ANALYSIS
Number of observations                                          644

SUMMARY OF DATA
     Number of missing data patterns             6

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100
     PROPORTION OF DATA PRESENT
           Covariance Coverage
              R2            R3            R4
              ________      ________      ________
 R2             0.983
 R3             0.516         0.533
 R4             0.342         0.239         0.349


UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES
    R2
      Category 1    0.968          613.000
      Category 2    0.032           20.000
    R3
      Category 1    0.968          332.000
      Category 2    0.032           11.000
    R4
      Category 1    0.956          215.000
      Category 2    0.044           10.000

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION !!! Very good fit, despite parameters being fixed (auto-regression)
Number of Free Parameters                        6
Chi-Square Test of Model Fit
          Value                           1.606*
          Degrees of Freedom                   3
          P-Value                         0.6580
*   The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used
    for chi-square difference testing in the regular way.  MLM, MLR and WLSM
```

```
    chi-square difference testing is described on the Mplus website.  MLMV, WLSMV,
    and ULSMV difference testing is done using the DIFFTEST option.
RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.000
          90 Percent C.I.                    0.000  0.052
          Probability RMSEA <= .05           0.941
CFI/TLI
          CFI                                1.000
          TLI                                1.043
Chi-Square Test of Model Fit for the Baseline Model
          Value                             70.870
          Degrees of Freedom                     6
          P-Value                           0.0000
WRMR (Weighted Root Mean Square Residual)
          Value                              0.361

MODEL RESULTS

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

 R4       ON
    R3                 1.017      0.533      1.906      0.057

 R3       ON
    R2                 0.849      0.315      2.695      0.007

 R2       ON
    R1                 1.333      0.291      4.578      0.000

 Thresholds
    R2$1               2.056      0.118     17.426      0.000
    R3$1               2.460      0.417      5.895      0.000
    R4$1               2.886      1.006      2.870      0.004

QUALITY OF NUMERICAL RESULTS
     Condition Number for the Information Matrix              0.100E-01
       (ratio of smallest to largest eigenvalue)

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS
!!! Junior year rape is not significantly predicted directly or indirectly by previous years
!!! but all effects for freshman and sophomore years are significant.
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Effects from R1 to R4
  Total                1.150      0.700      1.643      0.100
  Total indirect       1.150      0.700      1.643      0.100
  Specific indirect
    R4
    R3
    R2
    R1                 1.150      0.700      1.643      0.100

Effects from R2 to R4
  Total                0.863      0.519      1.661      0.097
  Total indirect       0.863      0.519      1.661      0.097
  Specific indirect
    R4
    R3
    R2                 0.863      0.519      1.661      0.097

Effects from R1 to R3
  Total                1.131      0.458      2.470      0.014
  Total indirect       1.131      0.458      2.470      0.014
  Specific indirect
    R3
    R2
    R1                 1.131      0.458      2.470      0.014

     Beginning Time:  18:32:50
        Ending Time:  18:33:00
       Elapsed Time:  00:00:10

MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA  90066
Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com
Copyright (c) 1998-2015 Muthen & Muthen
```


1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
Raped pre-college
Did not rape pre-college
Raped Freshman year
Did not rape Freshman year
Raped Sophomore year
Did not rape Sophomore year
0.33
0.11
0.17
0.08
0.13
0.05
Freshman
Sophomore
Junior

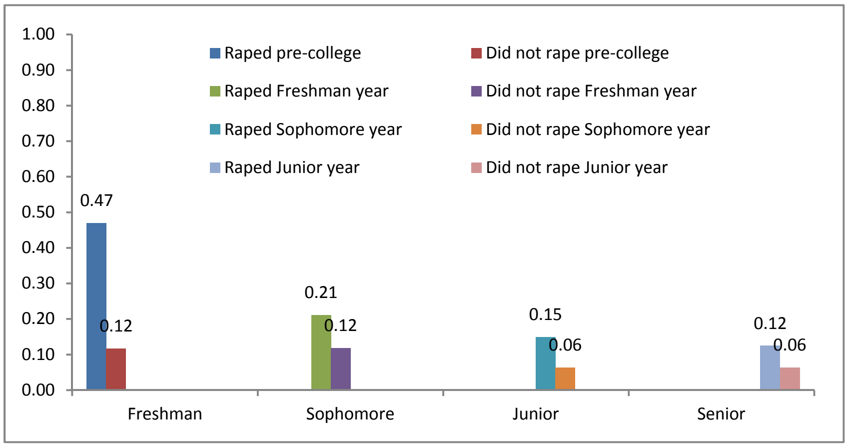
Appendix I: Autoregression (No Latent Classes) – Validation Dataset

```
Mplus VERSION 7.31
MUTHEN & MUTHEN
09/16/2015   7:20 PM

INPUT INSTRUCTIONS
  TITLE: Validation data - Autoregressive model;
  DATA: FILE IS RTraj.dat;
  VARIABLE: NAMES ARE id R1 R2 R3 R4 R5 dataset use;
      USEOBSERVATIONS ARE (dataset EQ 4);
      IDVARIABLE IS id;
      USEVARIABLES ARE R1-R5;
      CATEGORICAL ARE R2-R5;
      MISSING ARE all(9999999);
  ANALYSIS: TYPE = GENERAL; PARAMETERIZATION=THETA;
  MODEL:
      R5 ON R4;
      R4 ON R3;
      R3 ON R2;
      R2 ON R1;
  MODEL INDIRECT:
      R5 IND R1;
      R5 IND R2;
      R5 IND R3;
      R4 IND R1;
      R4 IND R2;
      R3 IND R1;

SUMMARY OF ANALYSIS
Number of observations                                         791

SUMMARY OF DATA
     Number of missing data patterns             8

COVARIANCE COVERAGE OF DATA
Minimum covariance coverage value   0.100

     PROPORTION OF DATA PRESENT

           Covariance Coverage
              R2            R3            R4            R5
              ________      ________      ________      ________
 R2             1.000
 R3             0.810         0.810
 R4             0.751         0.680         0.751
 R5             0.710         0.645         0.632         0.710

UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES
    R2
      Category 1    0.957          757.000
      Category 2    0.043           34.000
    R3
      Category 1    0.938          601.000
      Category 2    0.062           40.000
    R4
      Category 1    0.956          568.000
      Category 2    0.044           26.000
    R5
      Category 1    0.959          539.000
      Category 2    0.041           23.000

THE MODEL ESTIMATION TERMINATED NORMALLY

MODEL FIT INFORMATION
Number of Free Parameters                        8
Chi-Square Test of Model Fit
          Value                           9.716*  !!! Model fit is good.
          Degrees of Freedom                  6
          P-Value                        0.1371
*   The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used
    for chi-square difference testing in the regular way.  MLM, MLR and WLSM
    chi-square difference testing is described on the Mplus website.  MLMV, WLSMV,
    and ULSMV difference testing is done using the DIFFTEST option.
RMSEA (Root Mean Square Error Of Approximation)
          Estimate                        0.028  !!! Model fit is good.
          90 Percent C.I.                 0.000  0.059
          Probability RMSEA <= .05        0.865
CFI/TLI
          CFI                             0.984  !!! Model fit is good.
          TLI                             0.973
Chi-Square Test of Model Fit for the Baseline Model
```

```
          Value                              237.628
          Degrees of Freedom                      10
          P-Value                             0.0000
WRMR (Weighted Root Mean Square Residual)
          Value                                0.730

MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
 R5       ON
    R4                 0.750      0.207      3.626      0.000  !!! All autoregression
                                                               !!! parameters are significant.
 R4       ON
    R3                 0.939      0.220      4.275      0.000
 R3       ON
    R2                 0.698      0.159      4.376      0.000
 R2       ON
    R1                 1.906      0.222      8.581      0.000
 Thresholds
    R2$1               2.027      0.103     19.655      0.000
    R3$1               2.014      0.190     10.620      0.000
    R4$1               2.683      0.354      7.581      0.000
    R5$1               2.701      0.437      6.179      0.000

QUALITY OF NUMERICAL RESULTS
     Condition Number for the Information Matrix              0.109E-01
       (ratio of smallest to largest eigenvalue)

TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS
 !!! All indirect effects are significant.
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Effects from R1 to R5
  Total                0.936      0.331      2.824      0.005
  Total indirect       0.936      0.331      2.824      0.005
  Specific indirect
    R5
    R4
    R3
    R2
    R1                 0.936      0.331      2.824      0.005

Effects from R2 to R5
  Total                0.491      0.178      2.763      0.006
  Total indirect       0.491      0.178      2.763      0.006
  Specific indirect
    R5
    R4
    R3
    R2                 0.491      0.178      2.763      0.006

Effects from R3 to R5
  Total                0.704      0.216      3.256      0.001
  Total indirect       0.704      0.216      3.256      0.001
  Specific indirect
    R5
    R4
    R3                 0.704      0.216      3.256      0.001

Effects from R1 to R4
  Total                1.249      0.341      3.658      0.000
  Total indirect       1.249      0.341      3.658      0.000
  Specific indirect
    R4
    R3
    R2
    R1                 1.249      0.341      3.658      0.000

Effects from R2 to R4
  Total                0.655      0.178      3.690      0.000
  Total indirect       0.655      0.178      3.690      0.000
  Specific indirect
    R4
    R3
    R2                 0.655      0.178      3.690      0.000

Effects from R1 to R3
  Total                1.330      0.291      4.567      0.000
  Total indirect       1.330      0.291      4.567      0.000
  Specific indirect
    R3
    R2
    R1                 1.330      0.291      4.567      0.000
```

```
     Beginning Time:  19:20:46
        Ending Time:  19:20:46
       Elapsed Time:  00:00:00
```

MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA 90066
Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com



Appendix J: Reconstructing Analyses Variables From Original (Derivation) Dataset

```
*****************************************************************************************************.
*****************************************************************************************************.
****      Campus Rape Study
****      Prevalence of Serial Offenders
****      Analyst: Allison Tracy
****      Data Management_JAMA RECONSTRUCTION.sps
*****************************************************************************************************.
*****************************************************************************************************.

** September 06, 2015.
IMPORT
  FILE='03212-0002-Data.por'.
DATASET NAME JAMA_90_95 WINDOW=FRONT.

******************************************************************************************************
**  Parsing the missing and erroneously coded data.
******************************************************************************************************
*** Creating dummy indicators for dropout. This is based on the first 50 and last 50 variables.

COMPUTE DROPOUT2=NMISS(relstat2,marstat2,relinfl2,relatt2,fdsaprv2,mdsaprv2,tough2,great2,
emot2,brag2,busy2,center2,altrue2,rough2,helpful2,contest2,greed2,mean2,bossy2,noempat2,
indecis2,giveup2,notrust2,confide2,numone2,better2,revenge2,empath2,friendl2,fluster2,damage2,
goods2,lied2,weapon2,stelfam2,hit2,rowdy2,avodpay2,drnkpub2,steal2,change2,cheatex2,deface2,
soldmj2,soldoth2,phyfigh2,plagiar2,badnerv2,tense2,anxious2,depresd2,loscntr2,posaff2,mpos2,
mneg2,fpos2,tradatt2,chivlat2,malevio2,
disaprv2,prat2,vrat2,vagg2,vva2,pagg2,vpa2,intoxic2,delinq2,emottie2,psystrs2,psywell2,over2,vuln2,
give2,out2,dominat2,submiss2,notsub2,conform2,domsub2,hedonsm2,love2,xconsen2,xpressp2,xauthsp2,
xforcsp2,xattemp2,xdrugat2,xpressi2,xauthsi2,xdrugsi2,xforcsi2,xsexact2,expgrp2,spend2,sxother2,
sxyou2,intox2,excited2,tease2).

COMPUTE DROPOUT3=NMISS(livesit,livepart,fratsor,frstdrnk,frstdrug,frstsex,jobpay,frnddie,fmlydie,
famlyjai,parljob,parsplit,acdmprob,aids,finance,caracdnt,losepart,fearcrme,abugrlfd,bealone,fail,timcnfl,
fmlyill,selfill,prntdrug,lostjob,drivrev,fearpreg,engbrokn,lostfrnd,abortion,brokeup,fight,arrested,propvctm,
violvctm,edustat3,gradpln3,relstat3,relinfl3,relatt3,relimp3,relself3,rellive3,reldeci3,fdsaprv3,mdsaprv3,
respect3,birthcn3,canget3,malevio3,disaprv3,prat3,vrat3,vagg3,vva3,pagg3,vpa3,intoxic3,delinq3,
emottie3,psystrs3,psywell3,over3,vuln3,give3,out3,dominat3,submiss3,notsub3,conform3,domsub3,
hedonsm3,love3,xconsen3,xpressp3,xauthsp3,xforcsp3,xattemp3,xdrugat3,xpressi3,xauthsi3,xdrugsi3,
xforcsi3,xsexact3,rrelate3,rbackgr3,ractivi3,rreputa3,raccept3,rsmart3,rcultur3,expgrp3,recrel3,recsocb3,
recsoca3,recrepu3,recaccp3,recsmar3,reccult3).

COMPUTE DROPOUT4=NMISS(edustat4,gradpln4,relstat4,relinfl4,relatt4,relimp4,relself4,rellive4,
reldeci4,fdsaprv4,mdsaprv4,respect4,birthcn4,canget4,hugmom4,kissmom4,remarks4,fondmom4,
jokes4,argsex4,forcmom4,playboy4,xrated4,gamble,party,letter,dangact,hhike,nopltrp,intexp,resauth,
excidat,enjyilg,damage4,goods4,lied4,weapon4,stelfam4,hit4,rowdy4,avodpay4,drnkpub4,steal4,
cheatex4,deface4,soldmj4,soldoth4,phyfigh4,plagiar4,wrngdam4,
osisters,ysisters,twsistrs,ostsists,ystsists,twstsist,mother,father,mdiscuss,
fdiscuss,mgotinfo,fgotinfo,mgothelp,fgothelp,minsult,finsult,msulked,fsulked,mstomp,fstomp,mcried,
fcried,mspite,fspite,mthrthit,fthrthit,mhitsome,fhitsome,mthruat,fthruat,mthrhit,fthrhit,mblklve,
fblklve,mpushed,fpushed,mslap,fslap,mkicked,fkicked,mhitat,fhitat,mbeat,fbeat,mweapon,fweapon,
museweap,fuseweap,mothweap,fothweap).

COMPUTE DROPOUT5=NMISS(edustat5,gradpln5,relstat5,relinfl5,relatt5,relimp5,relself5,rellive5,
reldeci5,fdsaprv5,mdsaprv5,respect5,birthcn5,canget5,hugmom5,kissmom5,remarks5,fondmom5,
jokes5,argsex5,forcmom5,playboy5,xrated5,damage5,goods5,lied5,weapon5,stelfam5,hit5,rowdy5,
avodpay5,drnkpub5,steal5,cheatex5,deface5,soldmj5,soldoth5,phyfigh5,plagiar5,wrngdam5,wrngstl5,
wrnghit5,wrngthf5,nolike5,nbetter5,nogood5,nrespec5,noproud5,failure5,useless5,
pagg5,vpa5,intoxic5,delinq5,emottie5,psystrs5,psywell5,over5,vuln5,give5,
out5,dominat5,submiss5,notsub5,conform5,domsub5,hedonsm5,love5,xconsen5,xpressp5,xauthsp5,xforcsp5,
xattemp5,xdrugat5,xpressi5,xauthsi5,xdrugsi5,xforcsi5,xsexact5,recrel5,recsocb5,recsoca5,recrepu5,
recaccp5,recsmar5,reccult5,fstrelat,fstsocbd,fstsocac,fstreput,fstaccpt,fstsmart,fstcultr,rrelate5,
rbackgr5,ractivi5,rreputa5,raccept5,rsmart5,rcultur5).

FORMATS DROPOUT2 DROPOUT3 DROPOUT4 DROPOUT5 (F8.0).

* For some reason, there are some values present even for those who dropped out so I had to use 95 as the
cutoff.
RECODE DROPOUT2 (LO THRU 95=0) (ELSE=1).
RECODE DROPOUT3 (LO THRU 95=0) (ELSE=1).
RECODE DROPOUT4 (LO THRU 95=0) (ELSE=1).
RECODE DROPOUT5 (LO THRU 95=0) (ELSE=1).
VARIABLE LABELS DROPOUT2 "Subject did not participate at T2 (spring of Freshman year)".
VARIABLE LABELS DROPOUT3 "Subject did not participate at T2 (spring of Sophomore year)".
VARIABLE LABELS DROPOUT4 "Subject did not participate at T2 (spring of Junior year)".
VARIABLE LABELS DROPOUT5 "Subject did not participate at T2 (spring of Senior year)".
VALUE LABELS DROPOUT2 DROPOUT3 DROPOUT4 DROPOUT5
0 "participated"
1 "did not participate".
```

```
COMPUTE PATTERN_DROPOUT=(DROPOUT2*.1)+(DROPOUT3*.01)+(DROPOUT4*.001)+(DROPOUT5*.0001).
FORMATS PATTERN_DROPOUT (F8.4).
EXECUTE.
FREQUENCIES PATTERN_DROPOUT.
```

PATTERN_DROPOUT

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.0000	170	20.0	20.0	20.0
	.0001	226	26.6	26.6	46.5
	.0011	111	13.0	13.0	59.6
	.0100	12	1.4	1.4	61.0
	.0101	51	6.0	6.0	67.0
	.0111	114	13.4	13.4	80.4
	.1001	1	.1	.1	80.5
	.1110	7	.8	.8	81.3
	.1111	159	18.7	18.7	100.0
	Total	851	100.0	100.0	

```
EXECUTE.

RECODE expgrp2 expgrp3 expgrp4 expgrp5 (ELSE=COPY) INTO expgrp2_dropout expgrp3_dropout expgrp4_dropout
expgrp5_dropout.
IF (DROPOUT2=1) expgrp2_dropout=7.
IF (DROPOUT3=1) expgrp3_dropout=7.
IF (DROPOUT4=1) expgrp4_dropout=7.
IF (DROPOUT5=1) expgrp5_dropout=7.
VALUE LABELS  expgrp2_dropout expgrp3_dropout expgrp4_dropout expgrp5_dropout
1 "No sexual experience"
2 "Consensual sexual contact"
3 "Unwanted sexual contact"
4 "Coercive sexual contact"
5 "Sexual abuse"
6 "Sexual assault"
7 "did not participate/dropped out"
9 "unexplained missingness".
MISSING VALUES  expgrp2_dropout expgrp3_dropout expgrp4_dropout expgrp5_dropout (7,8,9).
FORMATS expgrp2_dropout expgrp3_dropout expgrp4_dropout expgrp5_dropout (F8.0).
EXECUTE.

*** Swartout's code as per his email August 3, 2015.
IF (expgrp eq 6) r.1 = 1.
IF (expgrp lt 6) r.1 = 0.
EXECUTE.
IF (expgrp2 eq 6) r.2 = 1.
IF (expgrp2 lt 6) r.2 = 0.
EXECUTE.
IF (expgrp3 eq 6) r.3 = 1.
IF (expgrp3 lt 6) r.3 = 0.
EXECUTE.
IF (expgrp4 eq 6) r.4 = 1.
IF (expgrp4 lt 6) r.4 = 0.
EXECUTE.
IF (expgrp5 eq 6) r.5 = 1.
IF (expgrp5 lt 6) r.5 = 0.
EXECUTE.
VARIABLE LABELS r.1 "Swartout's code for converting existing expgrp variables to rape indicators".
VARIABLE LABELS r.2 "Swartout's code for converting existing expgrp variables to rape indicators".
VARIABLE LABELS r.3 "Swartout's code for converting existing expgrp variables to rape indicators".
VARIABLE LABELS r.4 "Swartout's code for converting existing expgrp variables to rape indicators".
VARIABLE LABELS r.5 "Swartout's code for converting existing expgrp variables to rape indicators".
RECODE r.1 r.2 r.3 r.4 r.5 (MISSING=9).
FORMATS  r.1 r.2 r.3 r.4 r.5 (F8.0).
EXECUTE.

* Locating sources of missing data in Swartout's rape indicators.
RECODE r.1 r.2 r.3 r.4 r.5 (ELSE=COPY) INTO R.1_missing R.2_missing R.3_missing R.4_missing R.5_missing.
FORMATS R.1_missing R.2_missing R.3_missing R.4_missing R.5_missing (F8.0).
IF (NMISS(expgrp)=1) R.1_missing=9.
IF (NMISS(expgrp2)=1) R.2_missing=9.
IF (NMISS(expgrp3)=1) R.3_missing=9.
IF (NMISS(expgrp4)=1) R.4_missing=9.
IF (NMISS(expgrp5)=1) R.5_missing=9.
IF (DROPOUT2=1) R.2_missing=7.
IF (DROPOUT3=1) R.3_missing=7.
IF (DROPOUT4=1) R.4_missing=7.
IF (DROPOUT5=1) R.5_missing=7.
VARIABLE LABELS R.1_missing "Swartout's coding - missing values assigned".
VARIABLE LABELS R.2_missing "Swartout's coding - missing values assigned".
VARIABLE LABELS R.3_missing "Swartout's coding - missing values assigned".
VARIABLE LABELS R.4_missing "Swartout's coding - missing values assigned".
VARIABLE LABELS R.5_missing "Swartout's coding - missing values assigned".
```

```
VALUE LABELS R.1_missing R.2_missing R.3_missing R.4_missing R.5_missing
0 "did not rape"
1 "raped"
7 "did not participate/dropped out"
9 "missing expgrp".
MISSING VALUES  R.1_missing R.2_missing R.3_missing R.4_missing R.5_missing (7,8,9).

**** Recreating the expgrp variable, using the code found in the online codebook.
** Correcting some errors and assigning missing data values to the items used to create expgrp.

* 395  who dropped out by T3 had "never" values for CONSENT3 & PRESSSP3.
DO IF (NMISS(AUTHSP3,FORCESP3,ATTEMPT3,DRUGATT3,PRESSSI3,AUTHSI3,DRUGSI3,FORCESI3,SEXACTS3)=9).
RECODE CONSENT3 PRESSSP3 (ELSE=9).
END IF.
* CONSENT3 had no assigned missing data values.
MISSING VALUES CONSENT3 (9).
DO IF (NMISS(CONSENT,PRESSSP,AUTHSP,FORCESP,ATTEMPT,DRUGATT,PRESSSI,AUTHSI,DRUGSI,FORCESI,SEXACTS)=11).
RECODE CONSENT PRESSSP AUTHSP FORCESP ATTEMPT DRUGATT PRESSSI AUTHSI DRUGSI FORCESI SEXACTS (ELSE=8).
END IF.
DO IF
(NMISS(CONSENT2,PRESSSP2,AUTHSP2,FORCESP2,ATTEMPT2,DRUGATT2,PRESSSI2,AUTHSI2,DRUGSI2,FORCESI2,SEXACTS2)=11).
RECODE CONSENT2 PRESSSP2 AUTHSP2 FORCESP2 ATTEMPT2 DRUGATT2 PRESSSI2 AUTHSI2 DRUGSI2 FORCESI2 SEXACTS2
(ELSE=8).
END IF.
DO IF
(NMISS(CONSENT3,PRESSSP3,AUTHSP3,FORCESP3,ATTEMPT3,DRUGATT3,PRESSSI3,AUTHSI3,DRUGSI3,FORCESI3,SEXACTS3)=11).
RECODE CONSENT3 PRESSSP3 AUTHSP3 FORCESP3 ATTEMPT3 DRUGATT3 PRESSSI3 AUTHSI3 DRUGSI3 FORCESI3 SEXACTS3
(ELSE=8).
END IF.
DO IF
(NMISS(CONSENT4,PRESSSP4,AUTHSP4,FORCESP4,ATTEMPT4,DRUGATT4,PRESSSI4,AUTHSI4,DRUGSI4,FORCESI4,SEXACTS4)=11).
RECODE CONSENT4 PRESSSP4 AUTHSP4 FORCESP4 ATTEMPT4 DRUGATT4 PRESSSI4 AUTHSI4 DRUGSI4 FORCESI4 SEXACTS4
(ELSE=8).
END IF.
DO IF
(NMISS(CONSENT5,PRESSSP5,AUTHSP5,FORCESP5,ATTEMPT5,DRUGATT5,PRESSSI5,AUTHSI5,DRUGSI5,FORCESI5,SEXACTS5)=11).
RECODE CONSENT5 PRESSSP5 AUTHSP5 FORCESP5 ATTEMPT5 DRUGATT5 PRESSSI5 AUTHSI5 DRUGSI5 FORCESI5 SEXACTS5
(ELSE=8).
END IF.
DO IF (DROPOUT2=1).
RECODE CONSENT2 PRESSSP2 AUTHSP2 FORCESP2 ATTEMPT2 DRUGATT2 PRESSSI2 AUTHSI2 DRUGSI2 FORCESI2 SEXACTS2
(ELSE=7).
END IF.
DO IF (DROPOUT3=1).
RECODE CONSENT3 PRESSSP3 AUTHSP3 FORCESP3 ATTEMPT3 DRUGATT3 PRESSSI3 AUTHSI3 DRUGSI3 FORCESI3 SEXACTS3
(ELSE=7).
END IF.
DO IF (DROPOUT4=1).
RECODE CONSENT4 PRESSSP4 AUTHSP4 FORCESP4 ATTEMPT4 DRUGATT4 PRESSSI4 AUTHSI4 DRUGSI4 FORCESI4 SEXACTS4
(ELSE=7).
END IF.
DO IF (DROPOUT5=1).
RECODE CONSENT5 PRESSSP5 AUTHSP5 FORCESP5 ATTEMPT5 DRUGATT5 PRESSSI5 AUTHSI5 DRUGSI5 FORCESI5 SEXACTS5
(ELSE=7).
END IF.
VALUE LABELS CONSENT PRESSSP AUTHSP FORCESP ATTEMPT DRUGATT PRESSSI AUTHSI DRUGSI FORCESI SEXACTS
CONSENT2 PRESSSP2 AUTHSP2 FORCESP2 ATTEMPT2 DRUGATT2 PRESSSI2 AUTHSI2 DRUGSI2 FORCESI2 SEXACTS2
CONSENT3 PRESSSP3 AUTHSP3 FORCESP3 ATTEMPT3 DRUGATT3 PRESSSI3 AUTHSI3 DRUGSI3 FORCESI3 SEXACTS3
CONSENT4 PRESSSP4 AUTHSP4 FORCESP4 ATTEMPT4 DRUGATT4 PRESSSI4 AUTHSI4 DRUGSI4 FORCESI4 SEXACTS4
CONSENT5 PRESSSP5 AUTHSP5 FORCESP5 ATTEMPT5 DRUGATT5 PRESSSI5 AUTHSI5 DRUGSI5 FORCESI5 SEXACTS5
1 "Never"
2 "At least once"
7 "did not participate/dropped out"
8 "participated but missing all Koss variables"
9 "unexplained missingness".
MISSING VALUES CONSENT PRESSSP AUTHSP FORCESP ATTEMPT DRUGATT PRESSSI AUTHSI DRUGSI FORCESI SEXACTS
CONSENT2 PRESSSP2 AUTHSP2 FORCESP2 ATTEMPT2 DRUGATT2 PRESSSI2 AUTHSI2 DRUGSI2 FORCESI2 SEXACTS2
CONSENT3 PRESSSP3 AUTHSP3 FORCESP3 ATTEMPT3 DRUGATT3 PRESSSI3 AUTHSI3 DRUGSI3 FORCESI3 SEXACTS3
CONSENT4 PRESSSP4 AUTHSP4 FORCESP4 ATTEMPT4 DRUGATT4 PRESSSI4 AUTHSI4 DRUGSI4 FORCESI4 SEXACTS4
CONSENT5 PRESSSP5 AUTHSP5 FORCESP5 ATTEMPT5 DRUGATT5 PRESSSI5 AUTHSI5 DRUGSI5 FORCESI5 SEXACTS5 (7,8,9).
* DRUGSI5 had 850 coded as missing because 1 "never" had been listed with 9 as a missing value.
MISSING VALUES DRUGSI5 (9).
```

```
CROSSTABS
  /TABLES=XDRUGSI BY drugsi
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		drugsi I: INTERCOURSE BECAUSE OF DRUGS		
		1 Never	2 At least once	Total
xdrugsi I: INTERCOURSE BECAUSE OF DRUGS	1 Never	810	0	810
	2 1	0	17	17
	3 2	0	8	8
	4 3-5	0	3	3
	5 >5	0	4	4
	9 missing all 3 Koss sexual assault variables	9	0	9
Total		819	32	851

*9 miscoded cases

```
CROSSTABS
  /TABLES=XFORCESI BY forcesi
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		forcesi I: INTERCOURSE BECAUSE OF FORCE		
		1 Never	2 At least once	Total
xforcesi I: INTERCOURSE BECAUSE OF FORCE	1 Never	822	0	822
	2 1	0	6	6
	3 2	0	5	5
	4 3-5	0	4	4
	5 >5	0	2	2
	8 missing this but not all 3 assault variables	3	0	3
	9 missing all 3 Koss sexual assault variables	9	0	9
Total		834	17	851

*12 miscoded cases

```
CROSSTABS
  /TABLES=XSEXACTS BY sexacts
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		sexacts I: SEXUAL ACTS BECAUSE OF FORCE		
		1 Never	2 At least once	Total
xsexacts I: SEXUAL ACTS BECAUSE OF FORCE	1 Never	812	0	812
	2 1	0	8	8
	3 2	0	9	9
	4 3-5	0	1	1
	5 >5	0	5	5
	8 missing this but not all 3 assault variables	7	0	7
	9 missing all 3 Koss sexual assault variables	9	0	9
Total		828	23	851

*16 miscoded cases

```
CROSSTABS
  /TABLES=XDRUGSI2 BY drugsi2
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		drugsi2 II: INTERCOURSE BECAUSE OF DRUGS				
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	Total
xdrugsi2 II: INTERCOURSE BECAUSE OF DRUGS	1 Never	628	0	0	1	629
	2 1	0	9	0	0	9
	3 2	0	4	0	0	4
	4 3-5	0	2	0	0	2
	5 >5	0	2	0	0	2
	7 did not participate/dropped out	0	0	167	0	167
	9 missing all 3 Koss sexual assault variables	1	0	0	37	38
Total		629	17	167	38	851

*2 miscoded cases

```
CROSSTABS
  /TABLES=XFORCSI2 BY forcesi2
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		forcesi2 II: INTERCOURSE BECAUSE OF FORCE					
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	9 unexplained missingness	
xforcsi2 II: INTERCOURSE BECAUSE OF FORCE	1 Never	633	0	0	1	0	634
	2 1	0	3	0	0	0	3
	3 2	0	2	0	0	0	2
	5 >5	0	1	0	0	0	1
	7 did not participate/dropped out	0	0	167	0	0	167
	8 missing this but not all 3 assault variables	3	0	0	0	3	6
	9 missing all 3 Koss sexual assault variables	1	0	0	37	0	38
Total		637	6	167	38	3	851

*5 miscoded cases

```
CROSSTABS
  /TABLES=XSEXACT2 BY sexacts2
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		sexacts2 II: SEXUAL ACTS BECAUSE OF FORCE					
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	9 unexplained missingness	Total
xsexact2 II: SEXUAL ACTS BECAUSE OF FORCE	1 Never	623	0	0	1	0	624
	2 1	0	3	0	0	0	3
	3 2	0	3	0	0	0	3
	4 3-5	0	1	0	0	0	1
	7 did not participate/dropped out	0	0	167	0	0	167
	8 missing this but not all 3 assault variables	11	0	0	0	4	15
	9 missing all 3 Koss sexual assault variables	1	0	0	37	0	38
Total		635	7	167	38	4	851

*13 miscoded cases

```
CROSSTABS
  /TABLES=XDRUGSI3 BY drugsi3
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		drugsi3 III: INTERCOURSE BECAUSE OF DRUGS				
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	Total
xdrugsi3 III: INTERCOURSE BECAUSE OF DRUGS	1 Never	445	0	0	1	446
	2 1	0	4	0	0	4
	3 2	0	1	0	0	1
	4 3-5	0	3	0	0	3
	5 >5	0	1	0	0	1
	7 did not participate/dropped out	0	0	343	0	343
	9 missing all 3 Koss sexual assault variables	2	0	0	51	53
Total		447	9	343	52	851

*3 miscoded cases

```
CROSSTABS
  /TABLES=XFORCSI3 BY forcesi3
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		forcesi3 III: INTERCOURSE BECAUSE OF FORCE					
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	9 unexplained missingness	
xforcsi3 III: INTERCOURSE BECAUSE OF FORCE	1 Never	445	0	0	1	0	446
	2 1	0	3	0	0	0	3
	3 2	0	1	0	0	0	1
	5 >5	0	3	0	0	0	3
	7 did not participate/dropped out	0	0	343	0	0	343
	8 missing this but not all 3 assault variables	0	0	0	0	2	2
	9 missing all 3 Koss sexual assault variables	2	0	0	51	0	53
Total		447	7	343	52	2	851

*3 miscoded cases

```
CROSSTABS
  /TABLES=XSEXACT3 BY sexacts3
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		sexacts3 III: SEXUAL ACTS BECAUSE OF FORCE					
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	9 unexplained missingness	
xsexact3 III: SEXUAL ACTS BECAUSE OF FORCE	1 Never	436	0	0	1	0	437
	2 1	0	4	0	0	0	4
	4 3-5	0	3	0	0	0	3
	7 did not participate/dropped out	0	0	343	0	0	343
	8 missing this but not all 3 assault variables	8	0	0	0	3	11
	9 missing all 3 Koss sexual assault variables	2	0	0	51	0	53
Total		446	7	343	52	3	851

*11 miscoded cases

```
CROSSTABS
  /TABLES=XDRUGSI4 BY drugsi4
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		drugsi4 IV: INTERCOURSE BECAUSE OF DRUGS 1 Never	2 At least once	7 did not participate/dropped out
xdrugsi4 IV: INTERCOURSE BECAUSE OF DRUGS	1 Never	286	0	1
	2 1	0	2	0
	3 2	0	7	0
	4 3-5	0	2	0
	5 >5	0	1	0
	7 did not participate/dropped out	0	0	390
	9 missing all 3 Koss sexual assault variables	2	0	0
Total		288	12	391

		drugsi4 IV: INTERCOURSE BECAUSE OF DRUGS 8 participated but missing all Koss variables	Total
xdrugsi4 IV: INTERCOURSE BECAUSE OF DRUGS	1 Never	0	287
	2 1	0	2
	3 2	0	7
	4 3-5	0	2
	5 >5	0	1
	7 did not participate/dropped out	0	390
	9 missing all 3 Koss sexual assault variables	160	162
Total		160	851

*3 miscoded cases

```
CROSSTABS
  /TABLES=XFORCSI4 BY forcesi4
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		forcesi4 IV: INTERCOURSE BECAUSE OF FORCE 1 Never	2 At least once	7 did not participate/dropped out
xforcsi4 IV: INTERCOURSE BECAUSE OF FORCE	1 Never	290	0	1
	2 1	0	1	0
	3 2	0	3	0
	5 >5	0	4	0
	7 did not participate/dropped out	0	0	390
	9 missing all 3 Koss sexual assault variables	2	0	0
Total		292	8	391

		forcesi4 IV: INTERCOURSE BECAUSE OF FORCE 8 participated but missing all Koss variables	Total
xforcsi4 IV: INTERCOURSE BECAUSE OF FORCE	1 Never	0	291
	2 1	0	1
	3 2	0	3
	5 >5	0	4
	7 did not participate/dropped out	0	390
	9 missing all 3 Koss sexual assault variables	160	162
Total		160	851

*3 miscoded cases

```
CROSSTABS
  /TABLES=XSEXACT4 BY sexacts4
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		sexacts4 IV: SEXUAL ACTS BECAUSE OF FORCE				
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	Total
xsexact4 IV: SEXUAL ACTS BECAUSE OF FORCE	1 Never	287	0	0	0	287
	2 1	0	1	0	0	1
	4 3-5	0	4	0	0	4
	5 >5	0	3	0	0	3
	7 did not participate/dropped out	0	0	391	0	391
	8 missing this but not all 3 assault variables	3	0	0	0	3
	9 missing all 3 Koss sexual assault variables	2	0	0	160	162
Total		292	8	391	160	851

*5 miscoded cases

```
CROSSTABS
  /TABLES=XDRUGSI5 BY drugsi5
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		drugsi5 V: INTERCOURSE BECAUSE OF DRUGS				
		1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	Total
xdrugsi5 V: INTERCOURSE BECAUSE OF DRUGS	1 Never	144	0	0	0	144
	3 2	0	1	0	0	1
	7 did not participate/dropped out	0	0	662	0	662
	9 missing all 3 Koss sexual assault variables	0	0	0	44	44
Total		144	1	662	44	851

*0 miscoded cases

```
CROSSTABS
  /TABLES=XFORCSI5 BY forcesi5
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		forcesi5 V: INTERCOURSE BECAUSE OF FORCE 1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	Total
xforcsi5 V: INTERCOURSE BECAUSE OF FORCE	1 Never	144	0	0	0	144
	3 2	0	1	0	0	1
	7 did not participate/dropped out	0	0	662	0	662
	9 missing all 3 Koss sexual assault variables	0	0	0	44	44
Total		144	1	662	44	851

*0 miscoded cases

```
CROSSTABS
  /TABLES=XSEXACT5 BY sexacts5
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ASIS.
```

		sexacts5 V: SEXUAL ACTS BECAUSE OF FORCE 1 Never	2 At least once	7 did not participate/dropped out	8 participated but missing all Koss variables	Total
xsexact5 V: SEXUAL ACTS BECAUSE OF FORCE	1 Never	144	0	0	0	144
	2 1	0	1	0	0	1
	7 did not participate/dropped out	0	0	662	0	662
	9 missing all 3 Koss sexual assault variables	0	0	0	44	44
Total		144	1	662	44	851

*0 miscoded cases

```
** EXPGRP (KOSS CATEGORIES)
* NO SEXUAL EXPERIENCE. (All "never")
IF (CONSENT5 EQ 1 AND PRESSSP5 EQ 1 AND AUTHSP5 EQ 1 AND FORCESP5 EQ 1
AND ATTEMPT5 EQ 1 AND DRUGATT5 EQ 1 AND PRESSSI5 EQ 1 AND AUTHSI5 EQ 1
AND DRUGSI5 EQ 1 AND FORCESI5 EQ 1 AND SEXACTS5 EQ 1)
EXPGRP5_REDO = 1.
* CONSENSUAL SEXUAL CONTACT. (Only consent is "at least once")
IF (CONSENT5 EQ 2 AND PRESSSP5 EQ 1 AND AUTHSP5 EQ 1 AND FORCESP5 EQ 1
AND ATTEMPT5 EQ 1 AND DRUGATT5 EQ 1 AND PRESSSI5 EQ 1 AND AUTHSI5 EQ 1
AND DRUGSI5 EQ 1 AND FORCESI5 EQ 1 AND SEXACTS5 EQ 1)
EXPGRP5_REDO = 2.
```

```
* UNWANTED SEXUAL CONTACT. (One or more sex play items is "at least once")
IF ((PRESSSP5 EQ 2 OR AUTHSP5 EQ 2 OR FORCESP5 EQ 2) AND ATTEMPT5 EQ 1
AND DRUGATT5 EQ 1 AND PRESSSI5 EQ 1 AND AUTHSI5 EQ 1 AND DRUGSI5 EQ 1
AND FORCESI5 EQ 1 AND SEXACTS5 EQ 1)
EXPGRP5_REDO = 3.
* COERCIVE SEXUAL CONTACT. (Intercourse because of pressure and/or authority)
IF ((PRESSSI5 EQ 2 OR AUTHSI5 EQ 2) AND DRUGSI5 EQ 1 AND FORCESI5 EQ 1
AND SEXACTS5 EQ 1)
EXPGRP5_REDO = 4.
* SEXUAL ABUSE. (One or both attempted unwanted intercourse is "at least once")
IF ((ATTEMPT5 EQ 2 OR DRUGATT5 EQ 2) AND PRESSSI5 EQ 1 AND AUTHSI5 EQ 1
AND DRUGSI5 EQ 1 AND FORCESI5 EQ 1 AND SEXACTS5 EQ 1)
EXPGRP5_REDO = 5.
* SEXUAL ASSAULT. (At least one item assessing Intercourse of sexual acts because of force is "at least once")
IF (DRUGSI5 EQ 2 OR FORCESI5 EQ 2 OR SEXACTS5 EQ 2) EXPGRP5_REDO = 6.
EXECUTE.
IF (NMISS(CONSENT,PRESSSP,AUTHSP,FORCESP,ATTEMPT,DRUGATT,PRESSSI,AUTHSI,DRUGSI,FORCESI,SEXACTS)= 11)
EXPGRP_REDO = 9.
IF (NMISS(CONSENT2,PRESSSP2,AUTHSP2,FORCESP2,ATTEMPT2,DRUGATT2,PRESSSI2,AUTHSI2,DRUGSI2,FORCESI2,SEXACTS2)= 11)
EXPGRP2_REDO = 9.
IF (NMISS(CONSENT3,PRESSSP3,AUTHSP3,FORCESP3,ATTEMPT3,DRUGATT3,PRESSSI3,AUTHSI3,DRUGSI3,FORCESI3,SEXACTS3)= 11)
EXPGRP3_REDO = 9.
IF (NMISS(CONSENT4,PRESSSP4,AUTHSP4,FORCESP4,ATTEMPT4,DRUGATT4,PRESSSI4,AUTHSI4,DRUGSI4,FORCESI4,SEXACTS4) =
11) EXPGRP4_REDO = 9.
IF (NMISS(CONSENT5,PRESSSP5,AUTHSP5,FORCESP5,ATTEMPT5,DRUGATT5,PRESSSI5,AUTHSI5,DRUGSI5,FORCESI5,SEXACTS5) =
11) EXPGRP5_REDO = 9.
IF (DROPOUT2=1) EXPGRP2_REDO=7.
IF (DROPOUT3=1) EXPGRP3_REDO=7.
IF (DROPOUT4=1) EXPGRP4_REDO=7.
IF (DROPOUT5=1) EXPGRP5_REDO=7.
** There were a few cases missing data because none of the above logical arguments applied, although they had
data to draw from.
RECODE EXPGRP_REDO EXPGRP2_REDO EXPGRP3_REDO EXPGRP4_REDO EXPGRP5_REDO (SYSMIS=8) (ELSE=COPY).
VALUE LABELS  EXPGRP_REDO EXPGRP2_REDO EXPGRP3_REDO EXPGRP4_REDO EXPGRP5_REDO
1 "No sexual experience"
2 "Consensual sexual contact"
3 "Unwanted sexual contact"
4 "Coercive sexual contact"
5 "Sexual abuse"
6 "Sexual assault"
7 "did not participate/dropped out"
8 "spot missingness and faulty logic in code"
9 "missing all Koss variables".
FORMATS EXPGRP_REDO EXPGRP2_REDO EXPGRP3_REDO EXPGRP4_REDO EXPGRP5_REDO (F8.0).
MISSING VALUES expgrp expgrp2 expgrp3 expgrp4 expgrp5 EXPGRP_REDO EXPGRP2_REDO EXPGRP3_REDO EXPGRP4_REDO
EXPGRP5_REDO (99).

*** Alignment of EXPGRP variable existing in the dataset (rows) and the EXPGRP variable I created (columns).
CROSSTABS
  /TABLES=expgrp BY EXPGRP_REDO
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		EXPGRP_REDO		
		1 No sexual experience	2 Consensual sexual contact	3 Unwanted sexual contact
expgrp I: SEXUAL AGGRESSION GROUP	1 No sexual experience	230	0	0
	2 Consensual sexual contact	0	431	0
	3 Unwanted sexual contact	0	0	91
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	1	0
Total		230	432	91

		EXPGRP_REDO		
		4 Coercive sexual contact	5 Sexual abuse	6 Sexual assault
expgrp I: SEXUAL AGGRESSION GROUP	1 No sexual experience	0	0	0
	2 Consensual sexual contact	0	0	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	46	0	0
	5 Sexual abuse	0	9	0
	6 Sexual assault	0	0	43
Total		46	9	43

```
* 1 misassigned case.
```

```
CROSSTABS
  /TABLES=expgrp2 BY EXPGRP2_REDO
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		EXPGRP2_REDO		
		1 No sexual experience	2 Consensual sexual contact	3 Unwanted sexual contact
expgrp2 II: SEXUAL AGGRESSION GROUP	1 No sexual experience	218	0	0
	2 Consensual sexual contact	0	334	0
	3 Unwanted sexual contact	0	0	37
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	0	0
	9 Missing	0	0	0
Total		218	334	37

		EXPGRP2_REDO		
		4 Coercive sexual contact	5 Sexual abuse	6 Sexual assault
expgrp2 II: SEXUAL AGGRESSION GROUP	1 No sexual experience	0	0	0
	2 Consensual sexual contact	0	0	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	24	0	0
	5 Sexual abuse	0	6	0
	6 Sexual assault	0	0	20
	9 Missing	0	0	0
Total		24	6	20

		EXPGRP2_REDO		
		7 did not participate/dropped out	8 spot missingness and faulty logic in code	9 missing all Koss variables
expgrp2 II: SEXUAL AGGRESSION GROUP	1 No sexual experience	0	4	1
	2 Consensual sexual contact	0	3	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	0	0
	9 Missing	167	0	37
Total		167	7	38

* 8 misassigned cases.

```
CROSSTABS
  /TABLES=expgrp3 BY EXPGRP3_REDO
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		EXPGRP3_REDO		
		1 No sexual experience	2 Consensual sexual contact	3 Unwanted sexual contact
expgrp3 SEXUAL AGGRESSION GROUP	1 No sexual experience	151	0	0
	2 Consensual sexual contact	0	248	0
	3 Unwanted sexual contact	0	2	23
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	0	0
	9 Missing	0	0	0
Total		151	250	23

		EXPGRP3_REDO		
		4 Coercive sexual contact	5 Sexual abuse	6 Sexual assault
expgrp3 SEXUAL AGGRESSION GROUP	1 No sexual experience	0	0	0
	2 Consensual sexual contact	0	0	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	13	0	0
	5 Sexual abuse	0	5	0
	6 Sexual assault	0	0	12
	9 Missing	0	0	0
Total		13	5	12

		EXPGRP3_REDO		
		7 did not participate/dropped out	8 spot missingness and faulty logic in code	9 missing all Koss variables
expgrp3 SEXUAL AGGRESSION GROUP	1 No sexual experience	0	0	0
	2 Consensual sexual contact	0	2	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	0	0
	9 Missing	343	0	52
Total		343	2	52

* 4 misassigned cases.

```
CROSSTABS
  /TABLES=expgrp4 BY EXPGRP4_REDO
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		EXPGRP4_REDO		
		1 No sexual experience	2 Consensual sexual contact	3 Unwanted sexual contact
expgrp4 IV: SEXUAL AGGRESSION GROUP	1 No sexual experience	92	0	0
	2 Consensual sexual contact	0	169	0
	3 Unwanted sexual contact	0	0	17
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	0	0
	9 Missing	0	0	0
Total		92	169	17

		EXPGRP4_REDO		
		4 Coercive sexual contact	5 Sexual abuse	6 Sexual assault
expgrp4 IV: SEXUAL AGGRESSION GROUP	1 No sexual experience	0	0	0
	2 Consensual sexual contact	0	0	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	5	0	0
	5 Sexual abuse	0	3	0
	6 Sexual assault	0	0	12
	9 Missing	0	0	0
Total		5	3	12

		EXPGRP4_REDO		
		7 did not participate/dropped out	8 spot missingness and faulty logic in code	9 missing all Koss variables
expgrp4 IV: SEXUAL AGGRESSION GROUP	1 No sexual experience	0	2	0
	2 Consensual sexual contact	0	0	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	0	0
	9 Missing	391	0	160
Total		391	2	160

* 2 misassigned cases.

```
CROSSTABS
  /TABLES=expgrp5 BY EXPGRP5_REDO
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		EXPGRP5_REDO		
		1 No sexual experience	2 Consensual sexual contact	3 Unwanted sexual contact
expgrp5 V: SEXUAL AGGRESSION GROUP	1 No sexual experience	45	0	0
	2 Consensual sexual contact	0	89	0
	3 Unwanted sexual contact	0	0	7
	4 Coercive sexual contact	0	0	0
	5 Sexual abuse	0	0	0
	6 Sexual assault	0	0	0
	9 Missing	0	0	0
Total		45	89	7

		EXPGRP5_REDO		
		4 Coercive sexual contact	5 Sexual abuse	6 Sexual assault
expgrp5 V: SEXUAL AGGRESSION GROUP	1 No sexual experience	0	0	0
	2 Consensual sexual contact	0	0	0
	3 Unwanted sexual contact	0	0	0
	4 Coercive sexual contact	2	0	0
	5 Sexual abuse	0	1	0
	6 Sexual assault	0	0	1
	9 Missing	0	0	0
Total		2	1	1

		EXPGRP5_REDO		
		7 did not participate/dropped out	8 spot missingness and faulty logic in code	
expgrp5 V: SEXUAL AGGRESSION GROUP	1 No sexual experience	0	0	45
	2 Consensual sexual contact	0	0	89
	3 Unwanted sexual contact	0	0	7
	4 Coercive sexual contact	0	0	2
	5 Sexual abuse	0	0	1
	6 Sexual assault	0	0	1
	9 Missing	662	44	706
Total		662	44	851

* Fully replicated after correcting DRUGSI5 missing data value.

```
MISSING VALUES expgrp expgrp2 expgrp3 expgrp4 expgrp5 EXPGRP_REDO EXPGRP2_REDO EXPGRP3_REDO EXPGRP4_REDO
EXPGRP5_REDO (7,8,9).
** Swartout's code but based on the reconstructed Koss variable.
IF (expgrp_redo eq 6) r_redo_expgrp.1 = 1.
if (expgrp_redo lt 6) r_redo_expgrp.1 = 0.
EXECUTE.
IF (expgrp2_redo eq 6) r_redo_expgrp.2 = 1.
if (expgrp2_redo lt 6) r_redo_expgrp.2 = 0.
EXECUTE.
IF (expgrp3_redo eq 6) r_redo_expgrp.3 = 1.
if (expgrp3_redo lt 6) r_redo_expgrp.3 = 0.
EXECUTE.
IF (expgrp4_redo eq 6) r_redo_expgrp.4 = 1.
if (expgrp4_redo lt 6) r_redo_expgrp.4 = 0.
EXECUTE.
IF (expgrp5_redo eq 6) r_redo_expgrp.5 = 1.
if (expgrp5_redo lt 6) r_redo_expgrp.5 = 0.
EXECUTE.
* Locating sources of missing data.
IF (NMISS(expgrp_redo)=1) r_redo_expgrp.1=8.
IF (NMISS(expgrp2_redo)=1) r_redo_expgrp.2=8.
IF (NMISS(expgrp3_redo)=1) r_redo_expgrp.3=8.
IF (NMISS(expgrp4_redo)=1) r_redo_expgrp.4=8.
IF (NMISS(expgrp5_redo)=1) r_redo_expgrp.5=8.
IF (NMISS(CONSENT,PRESSSP,AUTHSP,FORCESP,ATTEMPT,DRUGATT,PRESSSI,AUTHSI,DRUGSI,FORCESI,SEXACTS)= 11)
r_redo_expgrp.1 = 9.
IF (NMISS(CONSENT2,PRESSSP2,AUTHSP2,FORCESP2,ATTEMPT2,DRUGATT2,PRESSSI2,AUTHSI2,DRUGSI2,FORCESI2,SEXACTS2)= 11)
r_redo_expgrp.2 = 9.
IF (NMISS(CONSENT3,PRESSSP3,AUTHSP3,FORCESP3,ATTEMPT3,DRUGATT3,PRESSSI3,AUTHSI3,DRUGSI3,FORCESI3,SEXACTS3)= 11)
r_redo_expgrp.3 = 9.
IF (NMISS(CONSENT4,PRESSSP4,AUTHSP4,FORCESP4,ATTEMPT4,DRUGATT4,PRESSSI4,AUTHSI4,DRUGSI4,FORCESI4,SEXACTS4) =
11)  r_redo_expgrp.4 = 9.
IF (NMISS(CONSENT5,PRESSSP5,AUTHSP5,FORCESP5,ATTEMPT5,DRUGATT5,PRESSSI5,AUTHSI5,DRUGSI5,FORCESI5,SEXACTS5) =
11)  r_redo_expgrp.5 = 9.
IF (DROPOUT2=1) r_redo_expgrp.2=7.
IF (DROPOUT3=1) r_redo_expgrp.3=7.
IF (DROPOUT4=1) r_redo_expgrp.4=7.
IF (DROPOUT5=1) r_redo_expgrp.5=7.
VALUE LABELS r_redo_expgrp.1 r_redo_expgrp.2 r_redo_expgrp.3 r_redo_expgrp.4 r_redo_expgrp.5
0 "did not rape"
1 "raped"
7 "did not participate/dropped out"
8 "participated but missing expgrp_redo"
9 "missing all Koss variables".
MISSING VALUES  r_redo_expgrp.1 r_redo_expgrp.2 r_redo_expgrp.3 r_redo_expgrp.4 r_redo_expgrp.5 (7,8,9).
VARIABLE LABELS r_redo_expgrp.1 "Recreating expgrp and using Swartout's code for rape indicators".
VARIABLE LABELS r_redo_expgrp.2 "Recreating expgrp and using Swartout's code for rape indicators".
VARIABLE LABELS r_redo_expgrp.3 "Recreating expgrp and using Swartout's code for rape indicators".
VARIABLE LABELS r_redo_expgrp.4 "Recreating expgrp and using Swartout's code for rape indicators".
VARIABLE LABELS r_redo_expgrp.5 "Recreating expgrp and using Swartout's code for rape indicators".
FORMATS  r_redo_expgrp.1 r_redo_expgrp.2 r_redo_expgrp.3 r_redo_expgrp.4 r_redo_expgrp.5 (F8.0).
EXECUTE.

MISSING VALUES R.1_missing R.2_missing R.3_missing R.4_missing R.5_missing r_redo_expgrp.1 r_redo_expgrp.2
r_redo_expgrp.3 r_redo_expgrp.4 r_redo_expgrp.5 (7,8,9).

*** The logic of the missing data assignment is faulty. If any indicator is 1, then “rape” can be assigned.
*** If all indicators are 0, then “no rape” can be assigned. HOWEVER, if there is a mix of missing and 0,
*** true rape status is not know and a missing data value must be assigned.
** Swartout's code with corrected missing based on assault variables.
* Reassign missing to explore patterns and crosstabs.
MISSING VALUES DRUGSI FORCESI SEXACTS DRUGSI2 FORCESI2 SEXACTS2 DRUGSI3 FORCESI3 SEXACTS3
DRUGSI4 FORCESI4 SEXACTS4 DRUGSI5 FORCESI5 SEXACTS5 (99).
COMPUTE PATTERN_T1_ASSAULT=(DRUGSI*.1) + (FORCESI*.01) + (SEXACTS*.001).
COMPUTE PATTERN_T2_ASSAULT=(DRUGSI2*.1) + (FORCESI2*.01) + (SEXACTS2*.001).
COMPUTE PATTERN_T3_ASSAULT=(DRUGSI3*.1) + (FORCESI3*.01) + (SEXACTS3*.001).
COMPUTE PATTERN_T4_ASSAULT=(DRUGSI4*.1) + (FORCESI4*.01) + (SEXACTS4*.001).
COMPUTE PATTERN_T5_ASSAULT=(DRUGSI5*.1) + (FORCESI5*.01) + (SEXACTS5*.001).
FORMATS  PATTERN_T1_ASSAULT PATTERN_T2_ASSAULT PATTERN_T3_ASSAULT PATTERN_T4_ASSAULT PATTERN_T5_ASSAULT (F8.3).
EXECUTE.
MISSING VALUES DRUGSI FORCESI SEXACTS DRUGSI2 FORCESI2 SEXACTS2 DRUGSI3 FORCESI3 SEXACTS3
DRUGSI4 FORCESI4 SEXACTS4 DRUGSI5 FORCESI5 SEXACTS5 (7,8,9).
MISSING VALUES expgrp_redo expgrp2_redo expgrp3_redo expgrp4_redo expgrp5_redo (99).
IF (expgrp_redo NE 6) r_redo_expgrp_missing.1 = 0.
IF (NMISS(DRUGSI,FORCESI,SEXACTS)>0) r_redo_expgrp_missing.1 = 9.
IF (expgrp_redo EQ 8) r_redo_expgrp_missing.1 = 8.
IF (expgrp_redo EQ 6) r_redo_expgrp_missing.1 = 1.
EXECUTE.
IF (expgrp2_redo NE 6) r_redo_expgrp_missing.2 = 0.
IF (NMISS(DRUGSI2,FORCESI2,SEXACTS2)>0) r_redo_expgrp_missing.2 = 9.
IF (expgrp2_redo EQ 8) r_redo_expgrp_missing.2 = 8.
IF (expgrp2_redo EQ 6) r_redo_expgrp_missing.2 = 1.
EXECUTE.
IF (expgrp3_redo NE 6) r_redo_expgrp_missing.3 = 0.
```

```
IF (NMISS(DRUGSI3,FORCESI3,SEXACTS3)>0) r_redo_expgrp_missing.3 = 9.
IF (expgrp3_redo EQ 8) r_redo_expgrp_missing.3 = 8.
IF (expgrp3_redo EQ 6) r_redo_expgrp_missing.3 = 1.
EXECUTE.
IF (expgrp4_redo NE 6) r_redo_expgrp_missing.4 = 0.
IF (NMISS(DRUGSI4,FORCESI4,SEXACTS4)>0) r_redo_expgrp_missing.4 = 9.
IF (expgrp4_redo EQ 8) r_redo_expgrp_missing.4 = 8.
IF (expgrp4_redo EQ 6) r_redo_expgrp_missing.4 = 1.
EXECUTE.
IF (expgrp5_redo NE 6) r_redo_expgrp_missing.5 = 0.
IF (NMISS(DRUGSI5,FORCESI5,SEXACTS5)>0) r_redo_expgrp_missing.5 = 9.
IF (expgrp5_redo EQ 8) r_redo_expgrp_missing.5 = 8.
IF (expgrp5_redo EQ 6) r_redo_expgrp_missing.5 = 1.
EXECUTE.
FORMATS r_redo_expgrp_missing.5 (F8.0).
MISSING VALUES r_redo_expgrp.1 r_redo_expgrp.2 r_redo_expgrp.3 r_redo_expgrp.4 r_redo_expgrp.5 (7,8,9).
IF (DROPOUT2=1) r_redo_expgrp_missing.2=7.
IF (DROPOUT3=1) r_redo_expgrp_missing.3=7.
IF (DROPOUT4=1) r_redo_expgrp_missing.4=7.
IF (DROPOUT5=1) r_redo_expgrp_missing.5=7.
VALUE LABELS r_redo_expgrp_missing.1 r_redo_expgrp_missing.2 r_redo_expgrp_missing.3 r_redo_expgrp_missing.4
r_redo_expgrp_missing.5
0 "did not rape"
1 "raped"
7 "did not participate/dropped out"
8 "participated but missing expgrp_redo"
9 "missing 3 Koss sexual assault variables".
FORMATS  r_redo_expgrp_missing.1 r_redo_expgrp_missing.2 r_redo_expgrp_missing.3 r_redo_expgrp_missing.4
r_redo_expgrp_missing.5 (F8.0).
VARIABLE LABELS r_redo_expgrp_missing.1 "Recoded rape variables based on recreated expgrp, corrected for
missing data".
VARIABLE LABELS r_redo_expgrp_missing.2 "Recoded rape variables based on recreated expgrp, corrected for
missing data".
VARIABLE LABELS r_redo_expgrp_missing.3 "Recoded rape variables based on recreated expgrp, corrected for
missing data".
VARIABLE LABELS r_redo_expgrp_missing.4 "Recoded rape variables based on recreated expgrp, corrected for
missing data".
VARIABLE LABELS r_redo_expgrp_missing.5 "Recoded rape variables based on recreated expgrp, corrected for
missing data".
MISSING VALUES   r_redo_expgrp_missing.1 r_redo_expgrp_missing.2 r_redo_expgrp_missing.3
r_redo_expgrp_missing.4 r_redo_expgrp_missing.5 (7,8,9).
MISSING VALUES  r.1 r.2 r.3 r.4 r.5 r_redo_expgrp_missing.1 r_redo_expgrp_missing.2 r_redo_expgrp_missing.3
r_redo_expgrp_missing.4 r_redo_expgrp_missing.5 (7,8,9).

*** Recreating the relevant category of the EXPGRP variable by directly using the original frequency variables.
*SEXUAL ASSAULT VICTIM.
RECODE XDRUGSI XFORCESI XSEXACTS XDRUGSI2 XFORCSI2 XSEXACT2
XDRUGSI3 XFORCSI3 XSEXACT3 XDRUGSI4 XFORCSI4 XSEXACT4 XDRUGSI5 XFORCSI5 XSEXACT5 (9=99) (ELSE=COPY).
MISSING VALUES XDRUGSI XFORCESI XSEXACTS XDRUGSI2 XFORCSI2 XSEXACT2
XDRUGSI3 XFORCSI3 XSEXACT3 XDRUGSI4 XFORCSI4 XSEXACT4 XDRUGSI5 XFORCSI5 XSEXACT5 (99).
DO IF (DROPOUT2=1).
RECODE XDRUGSI2 XFORCSI2 XSEXACT2 (MISSING=7).
END IF.
DO IF (DROPOUT3=1).
RECODE XDRUGSI3 XFORCSI3 XSEXACT3 (MISSING=7).
END IF.
DO IF (DROPOUT4=1).
RECODE XDRUGSI4 XFORCSI4 XSEXACT4 (MISSING=7).
END IF.
DO IF (DROPOUT5=1).
RECODE XDRUGSI5 XFORCSI5 XSEXACT5 (MISSING=7).
END IF.
DO IF (NMISS(XDRUGSI,XFORCESI,XSEXACTS)=3).
RECODE XDRUGSI XFORCESI XSEXACTS (MISSING=9).
END IF.
DO IF (NMISS(XDRUGSI2,XFORCSI2,XSEXACT2)=3).
RECODE XDRUGSI2 XFORCSI2 XSEXACT2 (MISSING=9).
END IF.
DO IF (NMISS(XDRUGSI3,XFORCSI3,XSEXACT3)=3).
RECODE XDRUGSI3 XFORCSI3 XSEXACT3 (MISSING=9).
END IF.
DO IF (NMISS(XDRUGSI4,XFORCSI4,XSEXACT4)=3) .
RECODE XDRUGSI4 XFORCSI4 XSEXACT4 (MISSING=9).
END IF.
DO IF (NMISS(XDRUGSI5,XFORCSI5,XSEXACT5)=3).
RECODE XDRUGSI5 XFORCSI5 XSEXACT5 (MISSING=9).
END IF.
RECODE XDRUGSI XFORCESI XSEXACTS XDRUGSI2 XFORCSI2 XSEXACT2
XDRUGSI3 XFORCSI3 XSEXACT3 XDRUGSI4 XFORCSI4 XSEXACT4 XDRUGSI5 XFORCSI5 XSEXACT5 (MISSING=8).
EXECUTE.
VALUE LABELS  XDRUGSI XFORCESI XSEXACTS XDRUGSI2 XFORCSI2 XSEXACT2
XDRUGSI3 XFORCSI3 XSEXACT3 XDRUGSI4 XFORCSI4 XSEXACT4 XDRUGSI5 XFORCSI5 XSEXACT5
1 "Never"
2 "1"
```

```
3 "2"
4 "3-5"
5 ">5"
7 "did not participate/dropped out"
8 "missing this but not all 3 assault variables"
9 "missing all 3 Koss sexual assault variables".
MISSING VALUES XDRUGSI XFORCESI XSEXACTS XDRUGSI2 XFORCSI2 XSEXACT2
XDRUGSI3 XFORCSI3 XSEXACT3 XDRUGSI4 XFORCSI4 XSEXACT4 XDRUGSI5 XFORCSI5 XSEXACT5 (7,8,9).
EXECUTE.
IF (XDRUGSI EQ 1 AND XFORCESI EQ 1 AND XSEXACTS EQ 1) RX.1 = 0.
IF (XDRUGSI2 EQ 1 AND XFORCSI2 EQ 1 AND XSEXACT2 EQ 1) RX.2 = 0.
IF (XDRUGSI3 EQ 1 AND XFORCSI3 EQ 1 AND XSEXACT3 EQ 1) RX.3 = 0.
IF (XDRUGSI4 EQ 1 AND XFORCSI4 EQ 1 AND XSEXACT4 EQ 1) RX.4 = 0.
IF (XDRUGSI5 EQ 1 AND XFORCSI5 EQ 1 AND XSEXACT5 EQ 1) RX.5 = 0.
IF (XDRUGSI GT 1 OR XFORCESI GT 1 OR XSEXACTS GT 1) RX.1 = 1.
IF (XDRUGSI2 GT 1 OR XFORCSI2 GT 1 OR XSEXACT2 GT 1) RX.2 = 1.
IF (XDRUGSI3 GT 1 OR XFORCSI3 GT 1 OR XSEXACT3 GT 1) RX.3 = 1.
IF (XDRUGSI4 GT 1 OR XFORCSI4 GT 1 OR XSEXACT4 GT 1) RX.4 = 1.
IF (XDRUGSI5 GT 1 OR XFORCSI5 GT 1 OR XSEXACT5 GT 1) RX.5 = 1.
IF (NMISS(XDRUGSI,XFORCESI,XSEXACTS)>0) RX.1 = 8.
IF (NMISS(XDRUGSI2,XFORCSI2,XSEXACT2)>0) RX.2 = 8.
IF (NMISS(XDRUGSI3,XFORCSI3,XSEXACT3)>0) RX.3 = 8.
IF (NMISS(XDRUGSI4,XFORCSI4,XSEXACT4)>0) RX.4 = 8.
IF (NMISS(XDRUGSI5,XFORCSI5,XSEXACT5)>0) RX.5 = 8.
IF (NMISS(XDRUGSI,XFORCESI,XSEXACTS)=3) RX.1 = 9.
IF (NMISS(XDRUGSI2,XFORCSI2,XSEXACT2)=3) RX.2 = 9.
IF (NMISS(XDRUGSI3,XFORCSI3,XSEXACT3)=3) RX.3 = 9.
IF (NMISS(XDRUGSI4,XFORCSI4,XSEXACT4)=3) RX.4 = 9.
IF (NMISS(XDRUGSI5,XFORCSI5,XSEXACT5)=3) RX.5 = 9.
IF (DROPOUT2=1) RX.2=7.
IF (DROPOUT3=1) RX.3=7.
IF (DROPOUT4=1) RX.4=7.
IF (DROPOUT5=1) RX.5=7.
VALUE LABELS RX.1 RX.2 RX.3 RX.4 RX.5
0 "did not rape"
1 "raped"
7 "did not participate/dropped out"
8 "combination of missing and never"
9 "missing 3 Koss sexual assault variables".
FORMATS  RX.1 RX.2 RX.3 RX.4 RX.5 (F8.0).
VARIABLE LABELS RX.1 "Recoded rape variables based on frequency variables, corrected for missing data".
VARIABLE LABELS RX.2 "Recoded rape variables based on frequency variables, corrected for missing data".
VARIABLE LABELS RX.3 "Recoded rape variables based on frequency variables, corrected for missing data".
VARIABLE LABELS RX.4 "Recoded rape variables based on frequency variables, corrected for missing data".
VARIABLE LABELS RX.5 "Recoded rape variables based on frequency variables, corrected for missing data".
MISSING VALUES RX.1 RX.2 RX.3 RX.4 RX.5 (7,8,9).

*** Alignment of R variables using raw frequency data, correct logic and missing data (row) with R variables using the existing expgrp variable and Swartout's original syntax code (column)
CROSSTABS
  /TABLES=RX.1 BY r.1
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		r.1 Swartout's code for converting existing expgrp variables to rape indicators		Total
		0	1	
RX.1 Recoded rape variables based on frequency variables, corrected for missing data	0 did not rape	790	1	791
	1 raped	0	42	42
	8 combination of missing and never	8	1	9
	9 missing 3 Koss sexual assault variables	9	0	9
Total		807	44	851

```
* 19 misclassified.
```

```
CROSSTABS
  /TABLES=RX.2 BY r.2
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		r.2 Swartout's code for converting existing expgrp variables to rape indicators		
		0	1	9
RX.2 Recoded rape variables based on frequency variables, corrected for missing data	0 did not rape	610	0	1
	1 raped	0	19	0
	7 did not participate/dropped out	0	0	167
	8 combination of missing and never	15	1	0
	9 missing 3 Koss sexual assault variables	2	0	36
Total		627	20	204

* 19 misclassified.

```
CROSSTABS
  /TABLES=RX.3 BY r.3
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		r.3 Swartout's code for converting existing expgrp variables to rape indicators		
		0	1	9
RX.3 Recoded rape variables based on frequency variables, corrected for missing data	0 did not rape	432	0	1
	1 raped	0	11	0
	7 did not participate/dropped out	0	0	343
	8 combination of missing and never	10	1	0
	9 missing 3 Koss sexual assault variables	2	0	51
Total		444	12	395

* 14 misclassified.

```
CROSSTABS
  /TABLES=RX.4 BY r.4
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		r.4 Swartout's code for converting existing expgrp variables to rape indicators		
		0	1	9
RX.4 Recoded rape variables based on frequency variables, corrected for missing data	0 did not rape	283	0	0
	1 raped	0	12	0
	7 did not participate/dropped out	0	0	391
	8 combination of missing and never	3	0	0
	9 missing 3 Koss sexual assault variables	2	0	160
Total		288	12	551

* 5 misclassified.

```
CROSSTABS
  /TABLES=RX.5 BY r.5
  /FORMAT=AVALUE TABLES
  /CELLS=COUNT
  /COUNT ROUND CELL.
```

		r.5 Swartout's code for converting existing expgrp variables to rape indicators		
		0	1	9
RX.5 Recoded rape variables based on frequency variables, corrected for missing data	0 did not rape	144	0	0
	1 raped	0	1	0
	7 did not participate/dropped out	0	0	662
	9 missing 3 Koss sexual assault variables	0	0	44
Total		144	1	706