2012

Measurement Timing In Growth Mixture Modeling of Alcohol Trajectories

Anne M. Fairlie
University of Rhode Island

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Recommended Citation

Fairlie, Anne M., "Measurement Timing In Growth Mixture Modeling of Alcohol Trajectories" (2012). *Open Access Dissertations*. Paper 1094.
https://digitalcommons.uri.edu/oa_diss/1094

This Dissertation is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

BF724.3
T56
F 357
2012

MEASUREMENT TIMING IN
GROWTH MIXTURE MODELING
OF ALCOHOL TRAJECTORIES

BY

ANNE M. FAIRLIE

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2012

819601538

ABSTRACT

A more nuanced understanding of individuals' patterns of alcohol use during adolescence, a key developmental period for the onset of use, is of critical importance for refining preventive interventions in this population. To this end, growth mixture modeling (GMM) is a statistical technique that may be used to identify latent subgroups of individuals who exhibit distinct patterns of alcohol use over time. Decisions regarding the timing and interval of survey assessments are particularly challenging in the context of GMM. Latent subgroups exhibit different trends in alcohol use, and these trends must be adequately captured by the survey assessments. Accordingly, the specific aims of the current research were to investigate how measurement timing (i.e., timing and spacing of assessments) affected the identification of the latent subgroups with: (1) an applied study using alcohol data from the National Longitudinal Survey of Youth (NLSY) 1997 and (2) a Monte Carlo simulation study. Participants from the NLSY1997 were 15 and 16 years old at Wave 1 ($n = 2686$, 49.44% female). Alterations in measurement timing were examined using five different assessment configurations: all 12 waves, two-year intervals, uneven intervals, the first six waves, and the last seven waves. The outcome, the number of drinks consumed per month, was assessed at each of 12 waves that spanned 11.5 years. The results of the applied study with the NLSY data were used as population parameters in the simulation study. The experimental factors investigated in the simulation study were measurement timing and sample size. First, the applied study revealed that the five-class GMM results were very similar when using all 12 waves versus two-year intervals. Only four participants were misclassified (i.e., assigned to

subgroups with different average alcohol trajectories). Second, the five-class GMM results when comparing all 12 waves to either the configuration with uneven intervals or the first six waves showed some degree of discrepancy with approximately 14% of the sample being misclassified. Third, the largest discrepancy in the five-class GMM results was observed when comparing the 12 wave and last seven wave configurations with 62% of the sample being misclassified. The simulation study showed that the 95% coverage estimates of the parameters (i.e., factor means, factor variances, factor covariance) were greater than .90 for four of the five assessment configurations, with the exception being the last seven waves. Three of the five assessment configurations produced average estimates of the parameters that were close to the population values. There was less precision in the parameter estimates, as indicated by larger average standard error estimates, for the configurations using the first six waves and the last seven waves. Collectively, these findings strongly suggest that the developmental window under investigation (i.e., all 12 waves versus the first six or last seven waves) had the most substantial impact on the reliability and validity of the five-class GMM solution. The sensitivity of the GMM solution to the timing of the survey assessments (i.e., developmental window) suggests that the latent classes should not be interpreted as representing subgroups that are present in the population. Instead, the identification of latent subgroups is sensitive to variations in research design, which include, but may not be limited to, measurement timing. It is important to better understand how these complex statistical approaches may be artifactually influenced by variations in research design. It may then be possible to have more informed evaluations of how prevention and intervention programs can alter individuals' patterns of alcohol use.

ACKNOWLEDGMENTS

This dissertation was supported by Award Number F31AA020164 from the National Institute on Alcohol Abuse and Alcoholism. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institute on Alcohol Abuse and Alcoholism or the National Institutes of Health.

I extend my sincerest appreciation to my co-major professors, Mark Wood and Theodore Walls, for assisting and guiding me throughout my dissertation and graduate school career. I am forever grateful for their mentorship.

I offer a special thank you to Kristina Jackson and Brian Flaherty for serving as consultants on my F31. I thank my dissertation committee members, Lisa Weyandt and Katherine Branch, for their helpful feedback. I also thank Joseph Rossi and Barbara Newman for serving on my examining committee.

I thank the National Institutes of Health for providing me with predoctoral funding to support my dissertation research and graduate training through a National Research Service Award (Grant F31AA020164).

Finally, I thank my parents and two sisters for always supporting me throughout my education and believing in my ability to accomplish whatever I set out to do.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS.....	v
LIST OF TABLES	vi
LIST OF FIGURES	x
CHAPTER 1	1
INTRODUCTION	1
CHAPTER 2	29
APPLIED STUDY: METHOD.....	29
CHAPTER 3	38
APPLIED STUDY: RESULTS	38
CHAPTER 4	55
SIMULATION STUDY: METHOD	55
CHAPTER 5	61
SIMULATION STUDY: RESULTS	61
CHAPTER 6	66
DISCUSSION	66
APPENDICES	155
BIBLIOGRAPHY	160

LIST OF TABLES

TABLE	PAGE
Table 1. Participants' ages across the 12 waves.	85
Table 2. Assessment configurations for the applied study and simulation study.	86
Table 3. Follow-up rate by wave.....	87
Table 4. Total number of waves completed for each participant.	88
Table 5. Descriptive information for drinks per month (original scale).	89
Table 6. Correlation coefficients for drinks per month using the original scale (upper triangle) and the logged scale (lower triangle).....	90
Table 7. Descriptive information for drinks per month (logged scale).....	91
Table 8. Model selection indices for the latent growth models using 12 waves.....	92
Table 9. Estimated parameters for the best fitting latent growth model using 12 waves.	93
Table 10. Model selection indices for the GMMs using 12 waves.....	94
Table 11. Class counts and proportions based on the estimated GMM using 12 waves.	95
Table 12. Average posterior probabilities for most likely class membership (row) by latent class (column) for 4- and 5-class GMMs using 12 waves.	96
Table 13. Drinks per month at Wave 1 by most likely class from the 5-class GMM using 12 waves.	97
Table 14. Median drinks per month and heavy episodic drinking by class for the 5- class GMM using 12 waves.	98

TABLE	PAGE
Table 15. Demographic information by class for the 5-class GMM using 12 waves.	99
Table 16. Estimated parameters for the best fitting 5-class GMM using 12 waves.	100
Table 17. Class counts and proportions based on the estimated 5-class GMMs for all five assessment configurations.....	101
Table 18. Average posterior probabilities for most likely class membership (row) by latent class (column) for the 5-class GMMs for all five assessment configurations.	102
Table 19. Estimated parameters for the 5-class GMMs for the four alternative assessment configurations.....	103
Table 20. Cohen's kappa to compare class agreement across assessment configurations.....	104
Table 21. Contingency table showing class agreement between the 5-class GMMs using 12 waves and 2-year intervals.	105
Table 22. Contingency table showing class agreement between the 5-class GMMs using 12 waves and uneven intervals.	106
Table 23. Contingency table showing class agreement between the 5-class GMMs using 12 waves and the first six waves.	107
Table 24. Contingency table showing class agreement between the 5-class GMMs using 12 waves and the last seven waves.....	108
Table 25. Percentage of participants assigned to Classes 1 to 5 in the alternative assessment configurations (column) based on class membership using 12 waves (row).	109

TABLE	PAGE
Table 26. Simulation results for the Class 1 factor means across the five assessment configurations, N = 1500.	110
Table 27. Simulation results for the Class 2 factor means across the five assessment configurations, N = 1500.	111
Table 28. Simulation results for the Class 3 factor means across the five assessment configurations, N = 1500.	112
Table 29. Simulation results for the Class 4 factor means across the five assessment configurations, N = 1500.	113
Table 30. Simulation results for the Class 5 factor means across the five assessment configurations, N = 1500.	114
Table 31. Simulation results for the factor variance and covariance parameters across the five assessment configurations, N = 1500.....	115
Table 32. Simulation results for the residual variances of the observed indicators (Y1-Y12) across the five assessment configurations, N = 1500.	116
Table 33. Simulation results for the Class 1 factor means across the five assessment configurations, N = 1000.	118
Table 34. Simulation results for the Class 2 factor means across the five assessment configurations, N = 1000.	119
Table 35. Simulation results for the Class 3 factor means across the five assessment configurations, N = 1000.	120
Table 36. Simulation results for the Class 4 factor means across the five assessment configurations, N = 1000.	121

TABLE	PAGE
Table 37. Simulation results for the Class 5 factor means across the five assessment configurations, N = 1000.	122
Table 38. Simulation results for the factor variance and covariance parameters across the five assessment configurations, N = 1000.....	123
Table 39. Simulation results for the residual variances of the observed indicators (Y1-Y12) across the five assessment configurations, N = 1000.	124
Table 40. Simulation results for the Class 1 factor means across the five assessment configurations, N = 500.	126
Table 41. Simulation results for the Class 2 factor means across the five assessment configurations, N = 500.	127
Table 42. Simulation results for the Class 3 factor means across the five assessment configurations, N = 500.	128
Table 43. Simulation results for the Class 4 factor means across the five assessment configurations, N = 500.	129
Table 44. Simulation results for the Class 5 factor means across the five assessment configurations, N = 500.	130
Table 45. Simulation results for the factor variance and covariance parameters across the five assessment configurations, N = 500.....	131
Table 46. Simulation results for the residual variances of the observed indicators (Y1-Y12) across the five assessment configurations, N = 500.	132

LIST OF FIGURES

FIGURE	PAGE
Figure 1. Unconditional latent growth curve model.	134
Figure 2. Unconditional growth mixture model.....	135
Figure 3. Profile plots for a random sample of participants (n = 15).....	136
Figure 4. Average trajectory from the latent growth model using 12 waves (N = 2686).	137
Figure 5. Average trajectory for each latent class from the 5-class growth mixture model using 12 waves (N = 2686).	138
Figure 6. Estimated trajectories for each participant in Class 5 (n = 96) and the average trajectory across all participants in Class 5 using 12 waves.	139
Figure 7. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus 2-year intervals (N = 2685).	140
Figure 8. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus uneven intervals (N = 2685).....	141
Figure 9. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus the 1 st six waves (N = 2685).....	142
Figure 10. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus the last seven waves (N = 2505).	143
Figure 11. Percentage of participants assigned to Classes 1 to 5 in the last seven wave configuration (legend) based on class membership using 12 waves (x-axis).	144
Figure 12. Average trajectories from the latent growth models for all five assessment configurations.....	145

FIGURE	PAGE
Figure 13. Sample size used in 31 applied studies.....	146
Figure 14. Profile plots of the simulated data from Class 1, Replication 1 (n = 339).	147
Figure 15. Profile plots of the simulated data from Class 5, Replication 1 (n = 61).	148
Figure 16. Average parameter estimates for the mean of the intercept factor for Class 1 by assessment configuration and sample size.	149
Figure 17. Average parameter estimates for the mean of the linear factor for Class 1 by assessment configuration and sample size.	150
Figure 18. Average parameter estimates for the mean of the quadratic factor for Class 1 by assessment configuration and sample size.	151
Figure 19. Average parameter estimates for the mean of the cubic factor for Class 1 by assessment configuration and sample size.	152
Figure 20. Average parameter estimates for the variance of the linear factor by assessment configuration and sample size.	153
Figure 21. Average parameter estimates for the variance of the quadratic factor by assessment configuration and sample size.	154

CHAPTER 1

INTRODUCTION

Alcohol Use: Prevalence and Patterns

Prevalence of alcohol use and its negative consequences. Adolescence and young adulthood are important developmental periods with respect to the initiation, escalation, and maintenance of alcohol use. According to the Monitoring the Future survey, prevalence rates for past 30 day alcohol use and heavy episodic drinking increase steadily across grades eight, ten and twelve (Johnston, O'Malley, Bachman, & Schulenberg, 2011a). Specifically, 13.8% of eighth graders, 28.9% of tenth graders and 41.2% of twelfth graders reported using alcohol in the previous 30 days (i.e., more than just a few sips). In addition, among eighth, tenth and twelfth graders, 7.2%, 16.3% and 23.2%, respectively, reported engaging in heavy episodic drinking (i.e., consuming five or more drinks in a row in the last two weeks).

Prevalence rates for alcohol use increase as individuals enter young or emerging adulthood (see Arnett, 2005), which is also the time period when many individuals attend college. Specifically, full-time college students (men: 44.3%, women: 32.2%) have been found to have higher prevalence rates for heavy episodic drinking compared to same-age non-college peers (men: 36.4%, women: 21.1%) (Johnston, O'Malley, Bachman, & Schulenberg, 2011b). Across the life course, prevalence rates for past 30 day alcohol use and heavy episodic drinking increase from age 18 through age 23 or 24 and then progressively decrease through age 50, particularly for heavy episodic drinking (Johnston et al., 2011b).

In addition to having the highest prevalence rates for heavy episodic drinking, young adults have an increased risk for alcohol use disorders. Alcohol use disorders include the clinical diagnoses of alcohol abuse and alcohol dependence as defined in the Diagnostic and Statistical Manual of Mental Disorders IV-TR (American Psychiatric Association, 2000). Alcohol abuse is characterized by a maladaptive pattern of alcohol use that results in significant impairment in social, interpersonal, or legal contexts (American Psychiatric Association, 2000, p. 199). Alcohol dependence is characterized by tolerance to alcohol's intoxicating effects, withdrawal symptoms, or an inability to control drinking (American Psychiatric Association, 2000, p. 197). Notably, the peak prevalence rates for alcohol use disorders correspond to the developmental period of young adulthood. As reported by Grant and colleagues (2006), individuals 18 to 29 years old exhibited the highest 12-month prevalence rates for alcohol abuse (6.95%) and alcohol dependence (9.24%) followed by prevalence rates of 5.95% for abuse and 3.77% for dependence among individuals 30 to 44 years old.

As individuals approach adulthood, there is a general decline in heavy drinking patterns, which has been referred to as "maturing out" (Johnston et al., 2011a, 2011b; Littlefield & Sher, 2010; Sher, Grekin, & Williams, 2005). The maturing out effect has been associated with role incompatibility related to marriage and work responsibilities (Littlefield & Sher, 2010; Sher et al., 2005). More recently, it has been suggested that changes in personality factors (e.g., decreased impulsivity) may also play a role in the observed decline in alcohol misuse as individuals approach adulthood (Littlefield, Sher, & Wood, 2009). In summary, adolescence and young

adulthood are critical developmental periods for the initiation, escalation, and maintenance of alcohol use as well as the development of alcohol use disorders.

Alcohol use, particularly heavy episodic drinking, has been associated with myriad negative consequences among adolescents and young adults, including fighting or arguing, doing things later regretted, engaging in unplanned sexual activity, and sustaining injuries (Perkins, 2002; Wechsler, Lee, Kuo, & Lee, 2000; Windle & Windle, 2006). It is estimated that, in 2005, approximately 5,534 individuals aged 18-24 years old in the United States were killed due to an alcohol-related unintentional injury, most often in motor vehicle accidents, with an estimated 1,825 of those being college students (Hingson, Zha, & Weitzman, 2009). The high occurrence of alcohol-related negative consequences among adolescents and young adults is particularly concerning from a public health standpoint. It is estimated that underage alcohol sales in the United States cost \$61.9 billion in 2001, including medical costs, property and work losses, and loss attributed to quality of life (Miller, Levy, Spicer, & Taylor, 2006).

Individuals' trends in alcohol use over time. Prevalence rates of alcohol use and negative consequences provide important information about alcohol-related behaviors. However, alcohol-related behaviors are extremely heterogeneous as indicated by the substantial individual variability with respect to the amount of alcohol consumed in a single sitting, the frequency with which individuals consume alcohol, and the pattern of drinking that develops over time. Therefore, more in-depth examinations of the patterns of alcohol use that individuals exhibit are particularly informative with respect to the course of alcohol use (Colder, Campbell, Ruel,

Richardson, & Flay, 2002; Greenbaum & Dedrick, 2007; Greenbaum, Del Boca, Darkes, Wang, & Goldman, 2005). Central to the current research, examination of alcohol use trajectories provides a more detailed examination of how individuals' alcohol use may change over time. Notably, the average alcohol trajectory from a sample does not reflect the meaningful variations in individuals' patterns on the outcome over time (Molenaar & Campbell, 2009). For example, certain individuals may exhibit limited alcohol use during a given time period, while other individuals exhibit high use with fluctuation over time.

A growing body of literature has sought to identify subgroups of individuals who exhibit distinct trends in alcohol use over time in an effort to characterize the heterogeneity of alcohol use patterns across individuals and inform preventive interventions. Accordingly, researchers have increasingly used statistical techniques that identify subgroups with different longitudinal patterns on an outcome measure (e.g., alcohol), specifically growth mixture modeling (GMM) and latent class growth analysis (LCGA) (Chassin, Pitts, & Prost, 2002; Goudriaan, Grekin, & Sher, 2007; Hix-Small, Duncan, Duncan, & Okut, 2004; Li, Duncan, Duncan, & Acock, 2001). Four prototypic classes have been identified (Sher, Jackson, & Steinley, 2011): 1) a non-drinking or non-heavy drinking class, 2) a class characterized by increasing use over time, 3) a class characterized by decreasing use over time, and 4) a relatively stable heavier drinking class. Findings on the estimated prevalence of each class and the shapes of the average class trajectories vary to some degree with respect to sample characteristics (e.g., age, sample size; Goudriaan et al., 2007; Hill, White, Chung, Hawkins, & Catalano, 2000; Tucker, Orlando, Ellickson, 2003), research design

characteristics (e.g., measurement timing; Jackson & Sher, 2006; Goudriaan et al., 2007; Greenbaum et al., 2005), and the indicator of alcohol use examined (e.g., heavy episodic drinking versus weekly drinking; Colder et al., 2002; Jackson & Sher, 2005; 2008).

In summary, a normative developmental increase in alcohol use has been observed during the time periods of adolescence and young adulthood, coupled with substantial individual variability over time. Sophisticated statistical approaches, like GMM and LCGA, offer important opportunities to advance longitudinal research since these techniques model the heterogeneity and can identify a set of homogenous subgroups. It is important, however, to better understand how these complex statistical approaches may be artifactually influenced by variations in research designs. One important design factor is measurement timing, which describes the intervals between survey assessments (e.g., time lag). Specifically, alterations in measurement timing may impact the identification of the homogeneous subgroups such that the subgroups may not be directly comparable across different assessment configurations (Cohen, 1991; Gollob & Reichardt, 1991; Sher & Jackson, 2006; Sher, Jackson, & Steinley, 2011). Accordingly, the effect that measurement timing may have on the identification of latent subgroups is the focus of the current investigation. Next, a description of two statistical techniques that have been commonly used for subgroup identification, namely GMM and LCGA, will be presented.

Primary Advantages of Mixture Modeling Techniques

The primary attraction to mixture modeling techniques, like GMM and LCGA, in the alcohol literature is due to the widespread recognition that alcohol use behaviors

are extremely heterogeneous (Brown et al., 2008). The two primary types of heterogeneity relate to intraindividual differences and interindividual differences (Molenaar & Campbell, 2009). *Intraindividual differences* are within-person differences that reflect, for instance, changes in alcohol consumption over time for a given individual. An individual may engage in heavy alcohol use during adolescence but then drink considerably less alcohol during young adulthood. *Interindividual differences* reflect between-person differences, whereby individuals may differ in the amount of alcohol they consume at a given point in time (e.g., alcohol use among a random sample of 15 year olds). Taken together, GMM and LCGA address heterogeneity associated with interindividual differences in intraindividual change, which is a type of heterogeneity that is fundamental to longitudinal studies. *Interindividual differences in intraindividual change* extend between-person differences to longitudinal data and represent the degree to which individuals vary in their patterns on the outcome over time. For example, certain individuals may consistently abstain from alcohol use, while other individuals may increase or decrease their alcohol consumption over time. GMM and LCGA are capable of modeling this type of heterogeneity and then summarizing individuals' patterns of alcohol use over time into several (e.g., four or five) commonly occurring drinking patterns.

The appeal of mixture modeling techniques is further strengthened by the ability of these techniques to provide more nuanced tests of etiological theories regarding the factors associated with the emergence and progression of alcohol use behaviors (Chassin, Flora, & King, 2004; Colder et al., 2002; Greenbaum et al., 2005;

Hill et al., 2000). For example, covariates can be added to mixture models to determine whether hypothesized etiological factors differentially predict class membership. For instance, are individuals who report high levels of impulsivity or sensation-seeking at baseline more likely to belong to a class with consistently high alcohol use throughout adolescence rather than a class characterized by a modest increase in alcohol use over time? Mixture models can also be used to investigate whether class membership may be associated with negative outcomes later in life (i.e., a distal outcome), such as high school dropout or adult alcohol use disorders. Therefore, mixture models offer great flexibility for investigating predictions based on etiological theories and the distal consequences associated with high-risk patterns of alcohol use.

Issues Associated with Mixture Modeling Techniques

Despite the advantages of applying GMM or LCGA to the study of alcohol use, there are several theoretical and empirical issues that warrant attention. To appropriately apply any statistical technique, it is critical that the limitations of the technique are recognized and attended to in real-world applications, especially in the case of complex statistical models. The issues presented are central to mixture modeling and its application in the psychological literature. The last issue described, measurement timing, is the focus of the current investigation.

Person- versus variable-centered. Distinctions among sampling dimensions have been well-established, such that researchers may sample among persons, variables, or occasions over time (Cattell, 1952; Molenaar & Campbell, 2009). When describing GMM and LCGA techniques, researchers have made a distinction between

person-centered and variable-centered statistical techniques (also called person- and variable-oriented) (Bauer, 2007; Bergman & Trost, 2006; Muthén & Muthén, 2000). The terms “person-centered” and “variable-centered” underscore that statistical techniques vary with respect to which sampling dimension is the focus of the technique. Despite the longstanding recognition of different types of sampling dimensions (Cattell, 1952), ambiguity can be observed regarding how to define and distinguish between person- and variable-centered approaches (Bergman & Trost, 2006). It appears that this ambiguity, at least partially, stems from the fact that person-centered approaches distinguish among individuals on the basis of the variable dimension. Both Muthén and Muthén (2000) and Laursen and Hoff (2006) describe variable-centered approaches as those that examine relations among variables. However, Muthén and Muthén describe person-centered approaches as those that examine similarities and differences among individuals, whereas Laursen and Hoff describe person-centered approaches as those that categorize individuals based on the relations among variables. Importantly, the method by which individuals are grouped into categories is based on their pattern of responses on a set of variables. GMM and LCGA identify subgroups of individuals based on the individuals’ responses on the repeated-measurements of the outcome of interest. Notably, Muthén and Muthén describe GMM and LCGA as statistical techniques that integrate variable- and person-centered approaches. A clearer and key distinction may be that certain statistical techniques assume a homogeneous population, while others assume a heterogeneous population and seek to summarize or characterize that heterogeneity in different ways (see also Molenaar & Campbell, 2009).

Survey instruments. Alcohol use questionnaires typically assess the occurrence of drinking behaviors for a given time period (e.g., past two weeks, past 30 days). Consequently, trends in alcohol use may be more difficult to adequately measure for individuals who drink infrequently or irregularly. An infrequent pattern of alcohol consumption may introduce bias into the ability of mixture modeling to accurately identify the alcohol use subgroups that may exist in the population. Furthermore, alcohol use is more opportunistic in early adolescence (Stanley, Henry, & Swaim, 2011), and among older adolescents and young adults alcohol consumption has been shown to vary in conjunction with external contingencies such as the academic calendar, weekends, and holidays (Greenbaum et al., 2005).

Rationale for the model specifications. Mixture models are inherently complex and require a detailed model-building process. For instance, should variances be estimated for the intercept (i.e., initial status) and slope (i.e., rate of change) factors to reflect individual variability in initial status and rate of change? If so, should these variances be constrained to be equal across the latent classes? Should linear and quadratic slopes be estimated for all the latent classes or just a portion of them? Given this complexity, it is not surprising that mixture models are known to have issues with model convergence, improper solutions, and lengthy computation times (Jung & Wickrama, 2008). It is unclear the degree to which model specifications are driven by theory, issues with model convergence, decisions made during the model-building process (e.g., inspection of plots and graphs), or the desire to decrease model complexity. The degree to which model specifications should be determined by each of these factors is not well delineated. It is clear that a substantial amount of time

should be allotted to model-building and careful decision-making throughout the model-fitting process (Bauer, 2007; Jung & Wickrama, 2008). Recognizably, journal requirements related to space constraints do not allow for a detailed description of the model-building process. However, greater attention should be given to delineating the motivations for the model specifications in published reports, given that the model specifications may impact the study's findings. For example, Bauer (2007) points out that the covariance matrix of the repeated measures is comprised of a portion relating to the class mean differences and a portion relating to the within-class covariance structure. Bauer underscores the effect that model misspecifications may have on model estimation, such that misspecification of the within-class covariance matrix may lead to over-identification of latent classes to account for the covariance among the repeated measures.

Ambiguity in determining the number of latent classes. The procedures for determining the appropriate number of classes in GMM and LCGA are not straightforward. Indeed, there is an extensive literature describing the issues that may arise when making decisions about the appropriate number of classes. These decisions likely have a substantial impact on the study's conclusions given that, in practice, the goal of GMM and LCGA is to identify commonly occurring drinking patterns that are conceptually meaningful. Several of the issues and cautions related to this ambiguity are highlighted next.

First, theory is often cited as one of the criteria that should be used to determine the appropriate number of classes (Jung & Wickrama, 2008). Researchers should consider whether or not the etiological literature supports the existence of the

identified subgroups for a given outcome (e.g., behavior). Bauer (2007), however, argues that theoretical justification for discrete groups or distinct types is most often lacking; instead, Bauer suggests that differences among individuals generally reflect differences in magnitude of the behavior.

Second, model fit indices (e.g., Bayesian Information Criterion or BIC) may be used to empirically determine the appropriate number of classes. Model fit indices do not always agree on which model is empirically favored, thus introducing a level of ambiguity into the model selection process (e.g., a three-class or a four-class model) (Bauer & Curran, 2004; Flaherty, 2010; Nylund, Asparouhov, & Muthén, 2007). It has been recommended that model fit indices should be used in conjunction with other indicators (e.g., theoretical justification, parsimony) to determine the best fitting model (Bauer & Curran, 2003b; Jung & Wickrama, 2008; Muthén, 2003; Rindskopf, 2003).

Third, researchers have emphasized that the number of latent classes should not be decided based solely on the unconditional model (e.g., no covariates) (Huang, Brecht, Hara, & Hser, 2010; Li & Hser, 2011; Muthén, 2004). Instead, a conditional model, which includes covariates to predict class membership, should inform the decision on the appropriate number of classes. There are numerous covariates that may potentially be included based on the etiological literature (e.g., parental monitoring, parental limit setting, alcohol expectancies, social modeling among peers; Brown et al., 2008). This introduces the possibility that the number of classes chosen would differ based on the covariates that are included in the model, thus decreasing the generalizability of the identified subgroups.

Fourth, Bauer and colleagues have identified several factors that may result in the identification of spurious latent classes, including non-normal data, nonlinear relations among variables, mis-specified models, violations of model assumptions, and local solutions (Bauer, 2007; Bauer & Curran, 2003a, 2004; Hipp & Bauer, 2006). Each of these alternative explanations cast doubt on the types of conclusions that may be drawn from mixture models.

The ambiguity and uncertainty associated with determining the correct number of latent classes is particularly troubling given that this decision is fundamental to any study that applies mixture modeling techniques. Taken together, these concerns highlight the complexities of mixture modeling and suggest that inferring the existence of the subgroups identified may lead to incorrect inferences about the population (Bauer, 2007).

Latent classes: Reality or heuristic? It is debatable as to whether or not the latent classes identified with mixture modeling techniques are representative of subgroups that truly exist in the population. Several researchers have advised against the over-generalization and over-reification of the latent classes, thus cautioning against the interpretation of the latent classes as a true or valid representation of subgroups that exist in the population (Bauer, 2007; Jackson & Sher, 2005, 2006; Nagin & Tremblay 2005; Sher et al., 2011). In particular, methodological factors (e.g., measurement timing, index of alcohol use examined) can affect the trajectory shape and prevalence of each latent class (Eggleston, Laub, & Sampson, 2004; Jackson & Sher, 2005, 2006; Sher et al., 2011). Furthermore, Bauer (2007) and Nagin and Tremblay (2005) contend that the latent classes impose a structure on what is better

considered a continuum of change. Nagin and Tremblay (2005) emphasize the utility of mixture modeling techniques for summarizing information provided by the individual-level trajectories into a set of more easily interpretable subgroups, which serve as a “convenient statistical approximation” (p. 882). Therefore, researchers employing mixture modeling techniques should avoid interpreting the latent classes as subgroups that definitively exist in the population.

Measurement timing. The final issue to be discussed is that of measurement timing, which describes the intervals between survey assessments (e.g., time lag). The current research focused on the effect of measurement timing on the trajectories of the latent classes identified with GMM. First, decision-making regarding the timing and spacing of survey assessments will be discussed with respect to longitudinal studies in general, and then issues specific to measurement timing and mixture modeling will be described.

The ability to test a well-articulated research hypothesis is intricately linked to the use of an appropriate study design and statistical analysis (Collins & Graham, 2002; Lerner, Schwartz, & Phelps, 2009; Nesselroade & Ghisletta, 2003). In longitudinal studies, it is critical that the timing of the assessments is determined by a conceptual understanding of how a given phenomena develops over time (i.e., rate of change over time). For example, studies on alcohol use behavior among college students may fail to capture important developmental increases in alcohol use that occur during the transition to college if the assessments are timed to begin once students have already matriculated. In addition to the pattern of change exhibited by a given phenomena, the empirical question being addressed may necessitate wider or

narrower intervals between assessments (i.e., decisions regarding the time lag) (Cohen, 1991; Gollob & Reichardt, 1991). If, for instance, the empirical question focuses on the daily patterns of alcohol use in relation to self-control, then multiple assessments each day may be required (Muraven, Collins, Shiffman, & Paty, 2005). Alternatively, researchers examining patterns of change over the life-course may be able to sufficiently capture change in the outcome with assessments that are spaced further apart (Jacob, Koenig, Howell, Wood, & Haber, 2009). Under circumstances where the interval necessary to observe changes in the phenomena is uncertain, then it is advisable to use the smallest interval for which change may be evident (Cohen, 1991; Gollob & Reichardt, 1991).

In practice, two general patterns of assessment spacing are commonly encountered for longitudinal panel studies in the alcohol literature. First, assessments may occur at even intervals throughout the study period, often occurring every year or every two years (Hix-Small et al., 2004; Li, Barrera, Hops, & Fisher, 2002; Wu, Witkiewitz, McMahon, Dodge, & Conduct Problems Prevention Research Group, 2010). Second, assessments may be spaced at 1-year or 2-year intervals in the earlier portion of a study period followed by one or two longer-term follow-ups (Chassin et al., 2004; Hill et al., 2000; Tucker et al., 2003; White, Johnson, & Buyske, 2000). This observation calls into question the degree to which individuals' patterns of alcohol use may be adequately measured in longitudinal studies.

The issue of how to appropriately study a behavior that changes over time has a long history in developmental psychology (Widaman, 1991; Wohlwill, 1973). Notably, Wohlwill (1973) recognized that inter-individual differences in a behavior

may vary over time, as in the case of alcohol use trajectory subgroups. For example, individuals may exhibit a similar level of the behavior initially and then diverge in the occurrence of the behavior over time. Numerous researchers have emphasized the importance of the number and spacing of assessments to accurately measure how an outcome changes over time (Collins & Graham, 2002; Jackson & Sher, 2006; Nesselroade & Ghisletta, 2003; Willett, 1989). Collins and Graham (2002) provided a thorough discussion of measurement timing in the context of tobacco and other drug use, although the identification of trajectory subgroups was not a focus of their discussion.

Decisions about measurement timing (i.e., the timing and spacing of the assessments) are further complicated in applications of mixture modeling, because the subgroups may exhibit different trends in alcohol use over time. In order to appropriately design studies, researchers should have *a priori* theories about the subgroups that may exist in the population. The hypothesized patterns of alcohol use for each of the subgroups should be used to determine the proper spacing of the assessments. Issues may arise if the timing of the survey assessments does not coincide with the alcohol use trajectory of a subgroup that may exist in the population. For example, the peak of an alcohol trajectory for an adolescence-limited drinking class may occur in mid-adolescence, while the peak for a normative class may occur during the early twenties. Alternatively, if a group of adolescents exhibit a peak in alcohol consumption around age 15 but assessments are spaced two years apart, then this potentially meaningful peak in alcohol use may be undetectable. Consequently, the latent classes identified via mixture modeling may be an artifact of the way the

assessments were spaced rather than reflective of the subgroups that may exist in the population. Accordingly, determining the optimal spacing of the repeated-measurements presents unique challenges due to the need to consider the shapes of the alcohol trajectories for each of the subgroups.

To date, three applied studies have examined the effect of measurement timing in the context of mixture modeling techniques (Eggleston et al., 2004; Jackson & Sher, 2006; Sher et al., 2011). Two of these studies examined alcohol use data from prospective longitudinal studies (Jackson & Sher, 2006; Sher et al., 2011). Few Monte Carlo simulation studies focusing on the identification of trajectory subgroups have examined the issue of measurement timing with respect to the total number of repeated-measurements of the outcome (Brown, 2003; Tolvanen, 2007). The current research provides a much-needed detailed examination of how measurement timing affects the identification of latent subgroups with different patterns of alcohol use during adolescence and young adulthood. Before reviewing findings from the empirical literature, the statistical model will be described in order to allow for a more comprehensive understanding of the empirical findings.

Details of the Statistical Model

Preliminary information. The statistical techniques that may be used to identify subgroups with distinct alcohol use trajectories (i.e., GMM, LCGA) build upon the technique of latent growth modeling. Therefore, a brief overview of latent growth modeling is provided. Latent growth modeling is a longitudinal modeling technique that evaluates the repeated-measurements of an outcome over time for a sample of individuals (Bollen & Curran, 2006; Laird & Ware, 1982; Singer & Willett,

2003). Latent growth modeling identifies the average trend in the outcome observed for a sample. Change in the outcome over time is statistically modeled with latent growth parameters, including an intercept factor and one or more slope factors. The term “latent” denotes that the growth parameters are estimated via the statistical model rather than directly observed or measured. In Figure 1, the repeated-measurements of the outcome are denoted as y_{i1} , y_{i2} , and y_{i3} . Measurements of the outcome (e.g., alcohol use) were taken at three different times for each individual i . A trajectory equation is estimated for each individual in order to characterize the change in the outcome over time (Bollen & Curran, 2006). The trajectory equation for the unconditional model (i.e., no covariates) is:

$$y_{it} = \alpha_i + \lambda_t \beta_i + \varepsilon_{it}$$

where y_{it} represents the outcome value y for each individual i at each time t . The ε_{it} is the error term for each individual i at each time t . A random intercept α_i and random slope β_i are estimated for each individual i . The intercept factor α_i typically reflects the initial level of the outcome, although it can be parameterized to reflect other time points (e.g., end status). One or more slope factors β_i are used to model the rate of change of the outcome over time (e.g., a linear increase in alcohol use over time). Polynomial functions (e.g., linear, quadratic, cubic) are used to define the slope parameters and model the shape of the individual-level trajectories. Linear and quadratic trends are commonly examined for alcohol data. The λ_t contains the coding values for time, which are constant across all individuals (note that subscript i is not included for λ_t). The values chosen for λ_t are used to parameterize the intercept and slope factors. In Figure 1, the coding for λ_t includes 1's for the intercept factor to

reflect the initial status on the outcome. The values 0, 1, and 2 are used to code the slope factor in order to reflect a linear change in the outcome over time and equal spacing between the assessments.

The intercept and slope factors are considered “random” since these parameters are allowed to vary across individuals. The variability around the intercept and slope values is incorporated into the statistical model through the estimation of variance parameters. Therefore, the variances of the intercept and slope factors account for heterogeneity around the initial status of the outcome and rate of change in the outcome over time. In Figure 1, the intercept factor and linear slope factor are allowed to correlate, thus a covariance parameter is also estimated. The intercept and slope factors are *continuous latent variables* (denoted by circles in Figure 1), which may take on a continuous range of values.

The equations for the intercept and slope factors in the unconditional latent growth model (e.g., no covariates) are:

$$\alpha_i = \mu_\alpha + \zeta_{\alpha i}$$

$$\beta_i = \mu_\beta + \zeta_{\beta i}$$

The intercept α_i for each individual is composed of the mean of the intercept factor μ_α and a disturbance term $\zeta_{\alpha i}$ for each individual i . The disturbance term reflects the individual’s deviation from the mean value. Similarly, the slope β_i for each individual is composed of the mean of the slope factor μ_β and the disturbance term $\zeta_{\beta i}$ for each individual. Therefore, an average trend is estimated for the entire sample, and the individual trajectories are allowed to vary around the average trend. Covariates may be added to the latent growth model in which case a conditional model (i.e., with

covariates) would be estimated. Covariates are used to predict the intercept and slope factors. For instance, conditional latent growth models can be used to investigate whether high levels on a measure of friends' alcohol use at baseline are associated with a high initial level of alcohol use or increases in alcohol use over time.

Growth mixture modeling. Growth mixture modeling (GMM) and latent class growth analysis (LCGA) are mixture modeling techniques that may be used to identify latent subgroups of individuals who exhibit distinct patterns in an outcome over time, namely a mixture of subgroups with different trajectories (Arminger, Stein, & Wittenberg, 1999; Jones & Nagin, 2007; Muthén, 2004; Muthén & Asparouhov, 2009; Muthén & Muthén, 2000; Nagin, 1999; Pearson, 1894). GMM and LCGA build upon the statistical model underlying latent growth modeling through the addition of a *categorical latent variable* (Figure 2). The categorical latent variable is used to characterize the subgroups of individuals whose pattern on the outcome follow, for example, different alcohol use trajectories. The categorical latent variable c_i denotes class membership ($c = 1, 2, \dots$, up to K classes) for each individual.

A critical aspect of GMM and LCGA is that class membership is unobserved (i.e., not measured directly such as by participant report). The proportion of the sample corresponding to each class is estimated via the statistical model, given that the proportion of individuals belonging to each class is not directly observable. In addition, the likelihood that an individual belongs to a particular class is derived through the calculation of posterior probabilities. For each individual, the probability of class membership is calculated for each of the latent classes being estimated. For example, if a three-class GMM is estimated, each individual will receive three

posterior probabilities—one for each of the three latent classes. The posterior probabilities are based on the individual's responses on the repeated-measurements of the outcome y_{it} . Ideally, each individual will have a high posterior probability for one of the latent classes (i.e., close to 1) and low posterior probabilities for the other latent classes (i.e., close to 0), indicating greater precision in the class assignments (Jung & Wickrama, 2008; Ram & Grimm, 2009).

Growth mixture models are typically used when the patterns of the outcome over time are considered to be heterogeneous. Mixture models assume that individuals exhibit a similar pattern of alcohol consumption within each latent class and that these patterns of alcohol consumption differ across the classes. In the case of GMM and LCGA, the latent subgroups are distinguished from each other by differences in their mean initial status (i.e., class-specific mean for the intercept factor) and mean rate of change in the outcome (i.e., class-specific mean for the slope factor). A mean trajectory (i.e., average trend) is estimated for each latent class, and these mean trajectories differ across the classes. In contrast, latent growth modeling estimates a single mean trajectory for the entire sample. A *critical distinction* between LCGA and GMM is that LCGA does not allow for individual variation around the estimated growth parameters within each class (e.g., the variances of the intercept and slope factors are fixed at zero), while GMM does allow for within-class variation (Muthén & Muthén, 2000; Nagin, 1999). Therefore, LCGA can be considered a special case of the more general GMM.

Computer software. In practice, there are two computer software programs that researchers typically use to apply GMM and LCGA in the alcohol literature: SAS

(SAS Institute Inc., 2008) and Mplus (Muthén & Muthén, 1998-2010), which is used in the analyses presented here. Although alternative software programs may be used to estimate GMMs and LCGAs, such as OpenMx (Grimm, Ram, & Estabrook, 2010), Mplus and SAS are the two software programs that are primarily used by alcohol researchers and, thus, will be briefly described here.

In SAS, the procedure PROC TRAJ may be used to estimate LCGAs, but this procedure cannot estimate GMMs (Jones & Nagin, 2007; Nagin, 1999, 2005). PROC TRAJ can handle outcomes with the following types of distributions: normal and censored normal distributions, Poisson and zero-inflated Poisson distributions, and a Bernoulli distribution. Maximum likelihood estimation is used to estimate the model parameters, and the likelihood function varies based on the type of data being analyzed. Time-invariant covariates (i.e., predictors that were measured at a single time point) and time-varying covariates (i.e., predictors that were measured longitudinally) can also be included in the LCGA to determine, for instance, whether etiological factors predict membership into the latent classes. An intercept factor plus linear, quadratic, and cubic slope factors may be estimated. This procedure can also handle estimation of a joint model where latent classes are estimated simultaneously for each of two separate outcomes.

In Mplus, both LCGA and GMM may be estimated, since the user can specify whether or not the variances of the latent factors are fixed to zero within the latent classes, as is characteristic of LCGA (Muthén & Asparouhov, 2009; Muthén & Muthén, 1998-2010; Muthén & Shedden, 1999). Mplus can handle data with a variety of distributions, including normal and censored normal distributions, Poisson and

zero-inflated Poisson distributions, and a Bernoulli distribution. As with PROC TRAJ, maximum likelihood estimation is used to estimate the model parameters. Time-invariant covariates, time-varying covariates, and distal outcomes may be included in the model. Users may estimate an intercept factor plus linear, quadratic, and cubic slope factors. Unlike PROC TRAJ, Mplus allows the user to estimate polynomial functions higher than a cubic factor. In addition, the Mplus procedure assumes that missing data are missing at random (MAR), while PROC TRAJ can only accommodate data that are assumed to be missing completely at random (MCAR) (Nagin, 2005). Mplus can handle additional variations on the model that are not currently supported by the SAS procedure (e.g., multiple group analysis, sequential GMM). For instance, a sequential GMM may be estimated where, for example, an outcome is measured over a series of occasions and latent classes are estimated for the first and second halves of the series with the latent classes from each half being related to one another.

Assessing Reliability and Validity of Statistical Techniques

Two analytical approaches have been used to evaluate the performance of GMM and LCGA: 1) applied studies analyzing prospective longitudinal data and 2) Monte Carlo simulation studies. Studies analyzing prospective longitudinal data represent real-world applications of the statistical technique to assess its reliability (i.e., the ability of the model to produce the same class solution across different data sets). Monte Carlo simulation studies enable researchers to evaluate the validity (or accuracy) of a statistical technique using computer-generated data. Next, these two

approaches will be described with a discussion of the relevant empirical research on GMM and LCGA.

Analyses on prospective longitudinal data. Few studies have analyzed prospective longitudinal data to examine how methodological factors impact the identification of alcohol use subgroups (Eggleston et al., 2004; Hipp & Bauer, 2006; Jackson & Sher, 2005, 2006, 2008; Sher et al., 2011). Studies using prospective longitudinal data are uniquely suited to assess the reliability of the latent subgroups obtained and can demonstrate how the identification of the subgroups may be influenced by methodological factors. For example, using alcohol use data that spanned the period from age 18 to age 30, Hipp and Bauer (2006) demonstrated the possibility of local solutions in the estimation of GMM and LCGA, which may occur when the best loglikelihood value is not replicated. Local solutions suggest the presence of unreliability in the identification of the latent subgroups, such that the results may not be replicated across studies. The potential for local solutions can be minimized by varying the starting values of the estimated parameters in the estimation procedure. In a series of articles, Jackson and Sher analyzed prospective data from two college student samples. The authors examined how the alcohol use subgroups varied according to the alcohol-related behavior examined (Jackson & Sher, 2005), measurement timing (Jackson & Sher, 2006), and the cutoffs of heavy episodic drinking (Jackson & Sher, 2008).

Germane to the current research, Jackson and Sher (2006) used data from 489 college freshmen (51% with a family history of alcoholism) to examine the effect of measurement timing on the class solutions for two alcohol outcomes: heavy episodic

drinking and alcohol use disorders (abuse or dependence). Follow-up assessments occurred annually for three years (Year 2, 3, and 4) followed by two long-term follow-ups at Years 7 and 11. The reliability of the mixture modeling technique was assessed across four assessment configurations: all six waves, removing Year 1, removing Year 4, and removing Year 11. Results showed differences across the assessment configurations with respect to the shape of the average trajectories for each class and class size (i.e., proportion of the sample). For instance, a developmentally limited class for heavy episodic drinking represented 18% of the sample when using all six waves, but only 6% of the sample when year 11 was omitted. Misclassification across the assessment configurations (i.e., assigned to subgroups with different average alcohol trajectories) was, at least in part, related to the probability of class membership and study attrition. Participants who were classified into similar classes had higher probabilities of class membership in the six wave solution than participants who were classified into different classes. For a subset of the comparisons between all six waves and the three alternative configurations, participants who were classified into similar classes had less missing data across the study period than participants who were classified into different classes. Lastly, omitting Year 4, which occurred during the transition out of college, appeared to result in the largest differences across the assessment configurations. The transition out of college corresponds to a developmental period typified by decreases in alcohol use, particularly heavy drinking (Johnson et al., 2011b). Notably, Jackson and Sher highlighted the flexibility of Monte Carlo simulation studies to vary both the number and spacing of assessments in order

to investigate how measurement timing impacts the alcohol trajectories used to identify the latent subgroups.

Sher and colleagues (2011) also examined how variations in measurement timing may affect the reliability of the solution obtained from mixture modeling. In this study, 3058 incoming college students completed eight assessments with Wave 1 occurring during the summer prior to matriculation and the follow-up assessments occurring for seven consecutive college semesters. Eight assessment configurations were tested to examine the degree to which the four prototypical classes (consistently low, increasing, decreasing, and consistently high) were observed across the different configurations. LCGA was used to estimate models for a binary indicator of heavy episodic drinking in the past 30 days (yes or no). Results showed that there was a general tendency the four prototypical classes to emerge across the eight assessment configurations; however, the shape of the average class trajectories varied to some degree across the solutions. Sher and colleagues underscored the discontinuities across the solutions given that the results using Waves 1 to 4 and those from Waves 5 to 8 did not correspond to the solution from all eight waves. Notably, participants who exhibited an increasing or decreasing trajectory in either the Waves 1-4 or Waves 5-8 configurations were the least consistently classified into a similar group using all 8 waves. Of particular significance, the authors highlighted the time-bounded nature of the solutions given that the “parts” did not equal the “whole.”

The findings from Jackson and Sher (2006) and Sher and colleagues (2011) provided important preliminary work on how alterations in measurement timing can affect the reliability of the class solution for alcohol trajectories. It is important to

examine non-college student samples in order to test the generalizability of these findings. Accordingly, the empirical application in the current study examined the drinking patterns of 15 and 16 year olds recruited from a national sample. In addition, a Monte Carlo simulation study was conducted on the topic of measurement timing, and the relevant literature on Monte Carlo studies is reviewed next.

Monte Carlo simulation studies. The application of statistical techniques in empirical research is universal across disciplines as a means of advancing knowledge and refining theories. Monte Carlo simulation studies are instrumental to the development and evaluation of statistical techniques in that they use computer-generated data to empirically test the performance of the technique under systematically varied conditions (Gentle, 2003; Mooney, 1997; Robert & Casella, 2010). Common motivations for simulation studies include: 1) testing the robustness of an analytical method when assumptions have been violated (Bauer & Curran, 2003a; Fan, Felsevalyi, Sivo, & Keenan, 2002), 2) developing a new statistical technique (Tooze, Grunwald, & Jones, 2002), and evaluating factors that may affect statistical power (Brown, 2003). Simulation studies offer a unique ability to know the “true” population from which the data were computer-generated, since the researcher must specify the characteristics of the true population. After generating data from the true population, these data are analyzed to determine the extent to which the “true” population parameters are recovered under different experimental conditions (e.g., sample size, varying measurement intervals). Experimental conditions used in simulation studies vary greatly depending on the research questions addressed and the statistical analysis under investigation. It is essential that the researcher limit the

number of experimental conditions to a manageable number, and the experimental conditions must be critical to the research question. As a result, researchers must attain a balance between manageability and scientific relevance (Paxton, Curran, Bollen, Kirby, & Chen, 2001). The need to achieve this balance underscores why a well-articulated research hypothesis must be developed in conjunction with a well-planned design and analysis in order to adequately address the empirical question (Paxton et al., 2001; Skrondal, 2000).

Despite the increasing use of GMM and LCGA, a limited number of simulation studies have examined how methodological factors may affect the performance of these statistical techniques. Existing studies using GMM and LCGA have evaluated the effects of a range of factors, including sample size, proportion of missing data, reliability of the observed variables, similarity of the class trajectories, non-normally distributed outcomes, and starting values for the estimated parameters (Bauer & Curran, 2003a, 2004; Brown, 2003; Hipp & Bauer, 2006; Liu, 2008; Nylund et al., 2007; Tolvanen, 2007). No general consensus has yet emerged on the conditions under which GMM or LCGA optimally perform with respect to a wide range of methodological factors. Germane to the current research, Brown (2003) tested six levels for the number of repeated-measurements (i.e., 3, 4, 5, 6, 9, and 12 with equal intervals between waves) with a single 2-class population. Brown found that statistical power increased substantially from three to six waves, where it reached .82, and then statistical power increased less dramatically up to 12 waves (.94). The results for the accuracy of detecting the correct 2-class class solution followed a similar pattern to those for statistical power. Notably, Bauer and Curran (2004) underscore the need for

future research to delineate the limitations and strengths of models that include both latent continuous and latent categorical variables, as found in GMM and LCGA. Accordingly, it is critical that we further examine how methodological factors may impact the performance of these statistical techniques. The current research is designed to directly address this knowledge gap with respect to measurement timing.

Goals of the Current Study

Despite the increasing use of GMM and LCGA, few studies have examined how measurement timing may affect the identification of the latent subgroups. As reviewed earlier, decision-making about how to space assessments in longitudinal studies is particularly challenging in the context of GMM. The alcohol use trends of each of the latent subgroups must be adequately captured by the assessments. Accordingly, the specific aims of the current study were to evaluate measurement timing in the context of GMM through the use of: 1) applied analyses on prospective longitudinal alcohol data from the National Longitudinal Survey of Youth (NLSY) 1997, and 2) a Monte Carlo simulation study that evaluates the impact of both measurement timing and sample size.

CHAPTER 2

APPLIED STUDY: METHOD

Specific Aim 1

The current study evaluated the extent to which the timing and spacing of assessments affected the identification of latent subgroups via GMM and the subgroups' average alcohol use trajectories using alcohol data from the NLSY1997. The general hypothesis was that if a particular subgroup exhibited a marked change in alcohol use (e.g., sharp increase), then the assessments that correspond to when this change occurred must be included in the analysis; otherwise, the reliability of identifying that shape of the alcohol use trajectory subgroup diminishes. Support for this hypothesis would underscore the need to consider the shape of the alcohol trajectories for all of the hypothesized theoretical subgroups when determining measurement timing.

Participants

Alcohol use data from the publically-available NLSY1997 cohort were utilized. The NLSY1997 data are ideally suited to test Specific Aim 1, which addresses the effect of measurement timing on the reliable identification of alcohol subgroups using an applied data set. First, the time period of the NLSY1997 spans adolescence into young adulthood and thus covers the developmental period of greatest interest to the current research. Second, the sample was recruited as a large and nationally representative sample from the U.S. population of adolescents. Third, the NLSY1997 data set includes 12 assessments to allow for flexibility in the timing and spacing of the assessments.

Participants were recruited in 1997 and were between the ages of 12 and 16 years at baseline. The first follow-up occurred approximately 18 months post-baseline, and ten subsequent follow-ups occurred annually for a total of 12 waves of data. Given the variability in alcohol use among individuals between the ages of 12 and 16 years, only individuals in the initial, nationally representative sample who were 15 and 16 years old at baseline were included in the analyses, resulting in a sample of 2,686 participants. By selecting this restricted cohort, the sample excludes younger participants who predominantly reported no drinking in the past month at baseline (e.g., 97.83% of 12 year olds). Older participants were also excluded since there were proportionally fewer older participants in the sample compared to 15 and 16 year olds (i.e., 1404 participants were 15 years old; 1282 participants were 16 years old; 491 participants were 17 years old).

Approximately half (49.44%) of the participants in the final sample ($n = 2686$) were female. All participants were 15 or 16 years old at Wave 1 ($M = 15.94$, $SD = 0.57$), and participants ranged in age from 25 to 29 years old at Wave 12 ($M = 27.50$, $SD = 0.63$) (Table 1). The majority of the participants were White/Caucasian (71.84%), followed by Black/African American (16.41%), Asian/Pacific Islander (2.55%), American Indian (0.60%), and another race not specified (8.60%). Twenty-three participants did not report race. The majority of the participants were non-Hispanic (85.92%) with nine participants not reporting ethnicity. The participants' place of residence was distributed across the United States: Northeast (19.02%), North Central (26.28%), South (34.25%), and West (20.44%).

Measures

Demographics. Demographic information included sex, race, ethnicity, and place of residence according to region in the United States.

Alcohol use. Across the 12 waves, alcohol use was assessed with five items: two screening items, a quantity item, a frequency item, and a heavy episodic drinking item. The first screening item assessed whether or not participants had ever had a drink of an alcoholic beverage. Participants were instructed that a drink was considered “a can or bottle of beer, a glass of wine, a mixed drink, or a shot of liquor.” The second screening item assessed whether or not participants had a drink of an alcoholic beverage since the date of the last interview. Participants who did not endorse drinking on the screening items were treated as zeros on the quantity, frequency, and heavy episodic drinking items to reflect no drinking.

The quantity and frequency items were open-ended and assessed alcohol use in the past month. The frequency item assessed the number of days in the last 30 days that participants had one or more drinks of an alcoholic beverage. The quantity item assessed the number of drinks participants usually consumed on the days that they drank. The quantity and frequency items were multiplied together, resulting in an outcome variable that reflects the number of drinks consumed per month.

The heavy episodic drinking item was used to verify the distinctions among the classes in the GMMs. Using an open-ended response format, participants reported the number of days they had consumed five or more drinks on the same occasion in the past 30 days. “Occasion” was defined as at the same time or within hours of each other.

Procedure

The survey was conducted by the Bureau of Labor Statistics, U.S. Department of Labor. The initial sample was randomly selected from households to be nationally representative ($n = 6,748$), plus a supplemental sample was recruited that over-sampled for Hispanic or Latino and African American individuals ($n = 2,236$). Participants had to be 12 to 16 years old as of December 31, 1996 in order to be eligible to take part in the survey. The sampling frame included all household residents in this age range. Therefore, a portion of the participants resided in the same household and most commonly had a sibling relationship. Siblings were not excluded from the subsample used in the current analyses ($n = 154$ siblings) given that the aims of the research were methodological in nature, and it was important to maintain a large sample size.

Interviewers administered the survey via a computer-assisted personal interview system that helps to minimize data errors and inconsistencies in responding. The youth questionnaire was administered via a one hour interview. Survey topics included education, employment, family attitudes, and substance use. Questions on substance use and other sensitive topics (e.g., criminal activity, sexual behavior) were administered with audio computer-assisted self-interview technology. This technology allowed participants to enter their responses directly into a computer without the interviewer being aware of the responses. Data collected from the youth questionnaire are publically available and may be downloaded at <https://www.nlsinfo.org/investigator/pages/login.jsp>.

Assessment Configurations

Four factors were considered when determining the assessment configurations. First, the assessment configurations were constructed based on the prototypical study designs used in GMM and LCGA studies on alcohol. In applied studies the spacing of the assessments often follows two prototypical patterns: 1) relatively evenly spaced assessments throughout the study period or 2) more frequent assessments in the beginning followed by assessments that are farther apart. Second, the developmental window under investigation may influence the identifiable subgroups, such that the GMM solutions using the first half and second half of a series do not necessarily correspond to the solution from the entire series (Sher et al., 2011). Third, the assessment configurations were determined with respect to the types of alcohol subgroups identified using all 12 assessments (e.g., a subgroup with higher alcohol consumption during adolescence). Fourth, the number of assessments limits the types of estimable slope parameters (e.g., a cubic slope cannot be estimated with only four waves). Accordingly, the assessment configurations were designed based on considerations of the prototypical study designs, the developmental window being investigated, the shape of the subgroup trajectories, and the estimable slope parameters.

Based on these considerations, five assessment configurations were tested in the current study (Table 2): 1) *All 12 waves*: All 12 available waves were included in the analysis, Waves 1-12. 2) *Two-year intervals*: Only the odd numbered waves were included (Waves 1, 3, 5, 7, 9, and 11). Waves 1 and 3 were 2.5 years apart while all the other intervals between assessments were two years. 3) *Uneven intervals*: Assessments were spaced more closely in the beginning of the study period with

longer follow-up intervals in the latter portion of the study period, Waves 1-4, 8, and 12. 4) *First six waves*: Only the first six waves were included in the analysis, Waves 1-6. 5) *Last seven waves*: Only the last seven waves were included in the analysis, Waves 6-12. Targeted comparisons were conducted to contrast the results from the configuration with all 12 waves to the results of the four alternative configurations.

Overview of the Applied Analyses

The following four major procedural steps were taken to analyze the NLSY alcohol data based on procedures and recommendations outlined in the literature (Jackson & Sher, 2005; Jung & Wickrama, 2008; Ram & Grimm, 2009).

First, latent growth modeling was used for model-building to determine the best fitting trajectory shape over the 12 waves for the full sample. This step determines the functional form of the outcome over time and dictates the most appropriate set of slope parameters to include in the subsequent growth mixture models (e.g., linear and quadratic slope parameters or linear, quadratic, and cubic slope parameters). Model selection indices for determining the best fitting latent growth model included: Akaike information criterion (AIC), Bayesian information criterion (BIC), sample size adjusted BIC (aBIC), comparative fit index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), chi-square difference test, and variance parameters for the latent factors. Improvements in model fit were indicated by: decreasing values for AIC, BIC, aBIC, RMSEA (less than .05), and SRMR (less than .05); increasing values for CFI (greater than .95) and TLI (greater than .95); significant chi-square difference tests; and significant variance parameters for the latent factors.

Second, GMM was conducted using all 12 waves to determine the number of latent classes that provided the best fit to the data. To do this, a series of GMMs that estimated two through seven latent classes was tested. For all of the models, each latent class included an intercept factor and the same set of slope factors. The set of slope factors (e.g., linear, quadratic, and cubic) was determined by the preliminary model-building step. Model selection indices for determining the best fitting GMM included: AIC, BIC, aBIC; Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR-LRT), adjusted Lo-Mendell-Rubin likelihood ratio test (adjusted LRT) (Lo, Mendell, & Rubin, 2001), and entropy. Other indicators for determining the best fitting GMM included: average posterior probabilities, class counts and proportions based on the estimated model, and inspection of the plots for the average trajectories for each latent class. Improvements in model fit were indicated by: decreasing values for AIC, BIC, and aBIC; increasing values for entropy (greater than .80; Ram & Grimm, 2009); significant VLMR-LRT and adjusted LRT; high average posterior probabilities for belonging to the most likely class (close to 1.0) and low average posterior probabilities for belonging to another class (close to 0.0); greater than 1% of the sample in each class (Jung & Wickrama, 2008); and separation between the average trajectories for each class with meaningful differences between the classes (Ram & Grimm, 2009).

Third, to test Specific Aim 1, GMMs were estimated in accordance with the assessment configurations displayed in Table 2. To allow for direct comparisons across the assessment configurations, the same number of latent classes included in the best fitting GMM using all 12 waves was also used in the estimation of the GMM for each alternative assessment configuration. For example, if the five-class GMM

provided the best fit to the data using all 12 waves, a five-class model using two-year intervals (Waves 1, 3, 5, 7, 9, and 11) was used as the comparison model. Indicators for comparing GMM results across the assessment configurations included: inspection of the plots for the class-specific average trajectories, class counts and proportions based on the estimated model, average posterior probabilities, and Cohen's kappa (Cohen, 1960). The GMM results of each assessment configuration were also compared by creating contingency tables to compare the degree to which participants were classified into a comparable alcohol use subgroup (i.e., similar trajectory shape) across the assessment configurations. The indicators used here for comparing results across the assessment configurations coincide with the procedures utilized in similar studies examining the reliability of mixture modeling (Jackson & Sher, 2005, 2006, 2008).

Fourth, the latent growth models and GMM were re-fit to determine the best fitting GMM for each alternative assessment configuration, as done in a similar study examining the reliability of mixture modeling (Sher et al., 2011). The results of this step were compared to determine whether the best-fitting GMM using all 12 waves coincided with the best-fitting GMM for each assessment configuration (e.g., five-class GMM using all 12 waves and five-class GMM using the first six waves).

Descriptive statistics were conducted in SAS Version 9.2 (SAS Institute Inc., 2008). All latent growth modeling and growth mixture modeling analyses were performed in Mplus Version 6 (Muthén & Muthén, 1998-2010). The outcome was modeled as a continuous measure with a normal distribution. For all Mplus analyses, missing data were handled using maximum likelihood estimation with robust standard

errors and an accelerated EM algorithm (Schafer & Graham, 2002; Muthén & Asparouhov, 2009; Muthén & Muthén, 1998-2010). Missing data are assumed to be missing at random (MAR), such that the probability of missingness is not related to the missing values of the outcome.

CHAPTER 3

APPLIED STUDY: RESULTS

Attrition

Across the 11 follow-up assessments, the follow-up rate was highest at Wave 2 (92.07%) and lowest at Wave 9 (77.74%) (Table 3). The last wave occurred approximately 11.5 years post-baseline and the follow-up rate was 80.57%. The follow-up rates reflect both participant dropout and intermittent non-response, where participants may have been unavailable at a given wave but completed the interview at a later follow-up. Just over half of the participants (56.48%) completed all 12 waves, followed by 12.25% who completed 11 waves, 6.37% who completed ten waves, and 4.43% who completed nine waves (Table 4). The majority of the participants (79.52%) completed nine or more of the 12 waves.

Exploratory Data Analysis

The distributional properties of the quantity (drinks per drinking day) and frequency (days drank per month) items were examined prior to creating the outcome measure reflecting drinks consumed in the last month. The responses on the quantity item ranged from 0 drinks per drinking day to 99 drinks per drinking day. Responses greater than 30 drinks per drinking day were considered extreme (e.g., implausible). These extreme responses were re-coded with ranking, such that the first extreme response at a given wave was re-coded as 31 and the next extreme response was re-coded as 32. When multiple participants provided the same extreme response, the midpoint was used. For example, two participants with the lowest extreme response were re-coded with a value of 31.5, which is the midpoint between 31 and 32.

Nineteen or fewer participants provided these extreme responses at any particular wave, reflecting a very small proportion of the data. After re-coding, the quantity and frequency values were multiplied together to form the outcome measure reflecting drinks consumed in the last month.

At Wave 1, almost three-quarters (71%) of the participants reported consuming no drinks in the last month (Table 5). At Wave 12, almost a third (32%) of the participants reported consuming no drinks in the last month. Across all 12 waves, the number of drinks that participants reported consuming in the last month ranged substantially (from 0 to more than 1,000 drinks in the last month). Skew and kurtosis were extremely high at each wave. At Waves 1 and 2, the median number of drinks consumed per month was zero and then gradually increased to six drinks at Wave 6 before leveling off. The correlations among the 12 outcome measures were relatively low, ranging from .03 to .43 (Table 6). Profile plots for a random sample of 15 participants are shown in Figure 3 to depict the fluctuations observed in drinking over time.

A constant (1.0) was added to all responses, since it is impossible to take a natural log of zero. After log transforming all responses, the skew and kurtosis values were less than two and three, respectively, at each time point (Table 7). The correlations among the 12 log transformed outcome measures ranged from .15 to .68 (Table 6). As expected, correlations were generally higher for adjacent waves, and correlations decreased as the time lag increased (e.g., Waves 1 and 3 versus Waves 1 and 10). The log transformed outcome measures were used in all subsequent analyses.

Model-Building Using All 12 Waves

As noted, latent growth modeling was used as the first step in model-building in order to determine the functional form of the outcome over time. Specifically, all 12 waves were included in the analysis to establish the highest polynomial function that best represented the data for the full sample ($N = 2,686$). Five unconditional (i.e., no covariates) latent growth models were tested to evaluate the inclusion of the intercept factor, a linear slope factor, a quadratic slope factor, a cubic slope factor, and a quartic slope factor. Each subsequent model included an additional slope factor to determine whether the additional factor improved model fit. The intercept factor was parameterized to reflect initial status in the outcome at Wave 1.

The latent growth model that included a quartic slope factor did not converge. For the four remaining latent growth models that were tested, the addition of each subsequent slope factor improved model fit as indicated by: 1) increasing values for CFI and TLI, and 2) decreasing values for RMSEA, SRMR, AIC, BIC, and aBIC (Table 8). Chi-square difference tests showed a significant improvement in model fit when a linear slope factor was estimated along with the intercept factor [$\Delta\chi^2(3) = 2765.23, p < .001$], when a quadratic slope factor was added [$\Delta\chi^2(4) = 1886.46, p < .001$], and when a cubic slope factor was added [$\Delta\chi^2(5) = 384.82, p < .001$]. Given the improvement in the model fit indices and the significant chi-square different test, the latent growth model that estimated up to the cubic slope factor was retained.

The average trajectory ($N = 2686$) from the best fitting latent growth model is shown in Figure 4. Overall, participants reported low levels of drinking at Wave 1 (corresponding to an average age of 15.94 years) then drinking increased steadily through Wave 6 (corresponding to an average age of 21.59 years). After Wave 6,

drinking was relatively stable with a slight downward trend at Wave 12 (corresponding to an average age of 27.50 years). Estimated parameters for the best fitting latent growth model are shown in Table 9. The factor means and factor variances were significantly different from zero. The non-zero factor variances suggested that applying a growth mixture model to these data was appropriate given the heterogeneity in the latent factors. Based on these model-building results, the intercept factor, a linear slope factor, a quadratic slope factor, and a cubic slope factor were included for estimation in all subsequent growth mixture models.

Growth Mixture Modeling Using 12 Waves

Model specifications. A series of growth mixture models were estimated to determine the number of latent classes that best represented the data. Using all 12 waves, 2-class through 7-class growth mixture models were tested. For all the GMMs, the means of the four latent factors were allowed to vary across classes, thus the factor means were class-specific. The variances and covariances for the latent factors and the residual variances of the observed variables were constrained to be equal across latent classes. GMMs were also estimated where both the means and variances of the latent factors were free to vary across classes. However, these models failed to converge, which is a common issue with the estimation of mixture models (e.g., Jackson & Sher, 2006; Jung & Wickrama, 2008). Accordingly, in all the GMMs reported, the variances and covariances of the latent factors and the residual variances of the observed variables were not class-specific. Although not class-specific, the variances of the latent factors were estimated in the GMMs and thus were allowed to capture the

individual heterogeneity that may be observed in these parameters as is characteristic of GMMs.

All GMMs were estimated with 500 initial stage random sets of starting values and 50 final stage optimizations to avoid local solutions (Hipp & Bauer, 2006). GMMs were first estimated without any additional model constraints. When the best likelihood value was not replicated, then the random starting values and final stage optimizations were increased (e.g., 1500 and 150, respectively). When models had inadmissible or inappropriate solutions (e.g., negative variances, correlations among latent factors close to or greater than 1.0), standard procedures were used to re-estimate the models to obtain an admissible solution. These procedures included fixing a variance equal to .001 or zero.

The two-class through five-class GMMs using 12 waves estimated without imposing additional model constraints. For the six-class and seven-class GMMs using 12 waves, the residual variance of the Wave 1 measure of the outcome was fixed at zero. In addition, for the seven-class GMM using 12 waves, the variance of the cubic factor was fixed at zero.

Model selection. Model selection indices for the one-class through seven-class models are shown in Table 10. Entropy increases and AIC, BIC, and aBIC all decrease as the number of classes increases, suggesting that a higher number of classes may provide a better fit. The six-class and seven-class models were not considered viable since those models both had a class that included less than 1% of the sample (Table 11; Jung & Wickrama, 2008). Therefore, the four-class and five-class GMMs using all 12 waves were compared to determine which provided the best fit. Neither the four-

class nor the five-class model was favored based on entropy, class counts, and class proportions. The four-class GMM was favored according to the VLMR-LRT and adjusted LRT. The five-class GMM was favored based on the AIC, BIC, and aBIC.

When considering the average posterior probabilities for the four-class and five-class GMMs, the average posterior probabilities are interpreted in the following manner. Each participant receives a posterior probability for class membership associated with each class estimated (e.g., five posterior probabilities per participant in a five-class GMM), and these probabilities sum to 1.0 for each participant. Using average posterior probabilities from the five-class GMM as an example (bottom half of Table 12), consider the participants who were most likely to belong to Class 1 (i.e., highest posterior probability for Class 1). Among these participants, the average posterior probability for Class 1 equals .997, and the average posterior probability that these participants belonged to Class 2 equals .001. On average, participants had a high probability for belonging to their most likely class and low probabilities for belonging to alternative classes. The average posterior probabilities for the four-class GMM and five-class GMM were acceptable (i.e., close to 1.0 on the diagonal and close to 0.0 on the off-diagonals) (Table 12). Overall, the average posterior probabilities were somewhat better for the five-class GMM.

Given that model selection criteria showed that the five-class GMM seemed to provide a better fit than the four-class GMM, the observed values for Wave 1 were inspected for each class. Inspection of the Wave 1 values according to participants' most likely class provided additional evidence that there was sufficient class separation (Table 13). Specifically, the observed values at Wave 1 did not overlap

across the classes. For instance, all participants in Class 1 had an observed value of 0 for drinks per month at Wave 1, while participants in Class 2 had observed values that ranged from 0.69 to 1.39. The class-specific average trajectories, which were plotted using the estimated means across waves, also suggested that the five-class GMM provided adequate separation and meaningful differences among the classes, particularly across the first six waves (Figure 5). Taken together, results indicated that the five-class GMM was the best fitting model using all 12 waves. A description of the five classes is presented next, based on the class-specific average trajectories (Figure 5) and the median number of drinks consumed at each wave (Table 14).

Participants in Class 1 (*Normative Class* ($n = 1916$), 71.33%), the largest class, generally reported consuming no drinks per month at Wave 1 with steadily increasing use through Wave 6, followed by a period of relatively stable use (Figure 5). Across all 12 waves, participants in Class 1 reported a median of zero heavy drinking days in the last month (i.e., five or more drinks on the same occasion) and a relatively low median number of drinks per month (Table 14).

Participants in Class 2 (*Low-Increasing Class* ($n = 227$), 8.45%) reported consuming a relatively low number of drinks per month at Wave 1 and consumption gradually increased before leveling off. Across all 12 waves, participants in Class 2 reported a median of zero heavy drinking days in the last month and an increasing median number of drinks per month.

Participants in Class 3 (*Low-Steady Class* ($n = 241$), 8.97%) reported consuming a moderate number of drinks per month at Wave 1 and consumption increased before leveling off. Across all 12 waves, participants in Class 3 reported a

median of zero or one heavy drinking days in the last month and an increasing then decreasing median number of drinks per month.

Participants in Class 4 (*High-Slowly Decreasing Class* ($n = 206$), 7.67%) reported consuming a high number of drinks per month at Wave 1 and consumption slowly decreased over time. Across all 12 waves, participants in Class 4 reported a median of three heavy drinking days at Wave 1 then either one or two heavy drinking days at subsequent waves. The median number of drinks per month remained relatively high across the study period.

Participants in Class 5 (*Extreme-Sharply Decreasing Class* ($n = 96$), 3.57%) reported consuming an extremely high number of drinks per month at Wave 1 and a decrease in drinking over time. Across all 12 waves, participants in Class 5 reported a median of seven heavy drinking days at Wave 1 then one to three heavy drinking days at subsequent waves. The median number of drinks per month was extremely high at Wave 1 and remained relatively high through Wave 8.

Demographic information for the participants belonging to each of the five classes is presented in Table 15. The *Normative Class* and *Low-Steady Class* consisted of approximately equal proportions of males and females. The *Low-Increasing Class* consisted of a larger proportion of females compared to males (55.51%). In contrast, the two classes with the heaviest drinking at Wave 1 each consisted of a larger proportion of males compared to females (*High-Slowly Decreasing Class*, 57.77%; *Extreme-Sharply Decreasing Class*, 63.54%).

The *Normative Class* had the smallest proportion of White participants (67.18%) compared to the other four classes. The two classes with the heaviest

drinking at Wave 1 consisted of the largest proportions of White participants (*High-Slowly Decreasing Class*, 89.22%; *Extreme-Sharply Decreasing Class*, 87.50%). The average age of the participants at Wave 1 increased steadily from the lowest drinking class (*Normative Class*, $M = 15.91$, $SD = 0.57$) to the highest drinking class (*Extreme-Sharply Decreasing Class*, $M = 16.17$, $SD = 0.59$).

For the five-class GMM solution (Table 16), the estimated means of the latent factors vary across the classes, which can also be seen in the figure depicting the class-specific average trajectories (Figure 5). The within-class variance of the intercept factor was not significantly different from zero. In contrast, the within-class variances of the three slope factors (linear, quadratic, and cubic) were significantly different from zero, reflecting variability in the slope estimates for the participants in a given class. Three of the six within-class covariance parameters were significantly different from zero. A negative correlation was observed between the quadratic factor and both the linear and cubic factors, while a positive correlation was observed between the linear and cubic factors.

To evaluate the variability among participants within a class, the individual-level and average trajectories were estimated in SAS Version 9.2 for the participants in Class 5. As demonstrated in Figure 6, there was a high degree of variability among the shapes of the individual-level trajectories, even among participants who belonged to the same latent class as determined by the five-class GMM.

Comparisons across the Assessment Configurations

Model specifications for the alternative configurations. In order to obtain model convergence, the following model specifications were imposed when

estimating the GMMs. *Two-year intervals*: The two-class model estimated without imposing addition constraints. The variance of the intercept factor was fixed at zero for the three-class through seven-class GMMs. *Uneven intervals*: The variance of the cubic factor was fixed at zero for the two-class through seven-class GMMs. *First six waves*: The two-class model estimated without imposing addition constraints. The residual variance of the Wave 1 measure of the outcome was fixed at zero for the three-class through seven-class GMMs. *Last seven waves*: The two-class model estimated without imposing addition constraints. The variance of the intercept factor was fixed at zero for the three-class through seven-class GMMs.

Holding the number of classes constant. Results from the five-class GMM using 12 waves were compared to the five-class GMM results for each of the four alternative assessment configurations. Comparisons were based on estimated class trajectories, class counts and proportions, average posterior probabilities, the estimated parameters from the five-class GMMs, Cohen's kappa, and contingency tables.

Estimated class trajectories. Class-specific average trajectories, which were plotted using the estimated means across waves, were compared for the 12 wave GMM and each of the four alternative configurations.

Two-year intervals: The class-specific average trajectories for the five classes were comparable between the GMM using 12 waves and the GMM with two-year intervals (Figure 7).

Uneven intervals: The class-specific average trajectories for Class 1 (*Normative Class*) were comparable between the GMM using 12 waves and the GMM with uneven intervals (Figure 8). However, the estimated means at Wave 1 were

higher for the GMM with uneven intervals for Classes 2 through 4. In Class 5 (*Extreme-Sharply Decreasing Class*), the average trajectory dropped more sharply in the GMM with uneven intervals.

First six waves: When considering the average trajectory for the full sample, the first six waves corresponded to the period of increased consumption (Figure 4). Therefore, a model comparison was made specific to this developmental window. The class-specific average trajectories for Class 1 (*Normative Class*) were comparable between the GMM using 12 waves and the GMM using only the first six waves (Figure 9). However, the estimated means at Wave 1 for Classes 2 through 4 were higher for the GMM with only the first six waves. In addition the shapes of the average trajectories for Class 4 (*High-Slowly Decreasing Class*) and Class 5 (*Extreme-Sharply Decreasing Class*) were noticeably different between the GMM using 12 waves and the GMM with only the first six waves.

Last seven waves: When considering the average trajectory for the full sample, the last seven waves corresponded to the period of stable consumption (Figure 4). Therefore, a model comparison was made specific to this developmental window. The class-specific average trajectories and the class proportions were markedly different when comparing the GMM using 12 waves and the GMM using only the last seven waves (Figure 10). When using only the last seven waves, the class-specific trajectories no longer showed stable consumption, but instead showed trajectories with markedly different shapes across the five classes.

Class counts and proportions. Class counts and proportions were compared across all five assessment configurations (Table 17). Based on most likely class

membership, class counts were almost identical between the five-class GMMs using 12 waves and two-year intervals. Class counts were also very similar when comparing results using 12 waves to results using either uneven intervals or the first six waves. Notably, the class counts were almost identical for the uneven intervals and the first six waves. The largest differences in class counts were observed for the comparison between all 12 waves and the last seven waves. For example, the normative class comprised 71% of the sample using 12 waves but the comparable, low drinking class using the last seven waves comprised only 35% of the sample.

Average posterior probabilities. The results of the comparisons for the average posterior probabilities were similar to the results comparing the class counts across the five assessment configurations (Table 18). Average posterior probabilities were most similar across the 12 wave and two-year interval configurations. Results were slightly less similar between the 12 wave configuration and either the uneven intervals or first six waves configurations. Results were similar between the uneven and first six wave configurations. Finally, the greatest discrepancy was observed between the 12 wave versus last seven wave configurations. The average posterior probabilities from the last seven waves were less favorable than those from all 12 waves (i.e., generally farther from 1.0 on the diagonal and farther from 0.0 on the off-diagonals), indicating less precision in the class assignments when estimating the GMM with the last seven waves.

Estimated parameters from the GMMs. The estimated parameters from the five-class GMMs for the four alternative assessment configurations are shown in Table 19. Estimated parameters include the class-specific means of the latent factors

and the variances and covariances of the latent factors. For each class, the estimated means of the intercept factors are relatively similar across the five assessment configurations (see also Table 16). There is greater variability among the assessment configurations for the estimated means of the three slope factors (i.e., linear, quadratic, cubic). The estimated variance and covariance parameters are similar between the 12 wave and two-year configurations. However, there is substantial variability in the estimated variance and covariance parameters when comparing the 12 wave configuration to the configurations with uneven intervals, the first six waves, or the last seven waves.

Cohen's kappa. Cohen's kappa was used to assess the degree to which participants were assigned to a similar class across the assessment configurations (Cohen, 1960). According to guidelines proposed by Landis and Koch (1977), kappa values indicate: slight agreement (0.00 to 0.20), fair agreement (0.21 to 0.40), moderate agreement (0.41 to 0.60), substantial agreement (0.61 to 0.80), or almost perfect agreement (0.81 to 1.00). Based on participants' most likely class membership, there was nearly perfect agreement in class assignment between the 12 wave and 2-year interval configurations with kappa equal to .99 (Table 20). Agreement between the 12 wave configuration and both the uneven intervals and first six waves was also high with kappa equal to .69. Notably, the uneven interval configuration and first six waves had nearly perfect agreement in class assignment given a kappa of .99. Finally, there was only slight agreement between the 12 wave versus last seven wave configurations with kappa equal to .10. The low level of agreement in class assignment may be partially due to differences in class proportions between these two

configurations (Figure 10). For example, Class 1 comprises 71% of the sample for the 12 wave configuration, but only 35% of the sample for the last seven wave configuration. Given the substantial differences in class size, it is not surprising that agreement in class assignment is low.

Contingency tables. Four contingency tables were calculated to examine the consistency in class assignment between the 12 wave configuration and each of the four alternative configurations. Class assignment was determined by participants' most likely class based on posterior probabilities. The contingency tables compared the categorical variables for class assignment: Class 1 (Normative, lowest drinking class) through Class 5 (Extreme-Sharply Decreasing Class). This analysis shows the extent to which the participants were assigned to a similar drinking class across the assessment configurations. In Tables 21 to 24, the diagonal values represent the degree of congruence among corresponding classes (i.e., the participant was classified into Class 1 in each of the two assessment configurations being compared). The off-diagonal values represent the degree of misclassification or incongruence since non-corresponding classes are the basis of the comparison (e.g., Class 1 for 12 waves versus Class 5 for uneven intervals). Cell chi-squares values indicate whether the cell frequencies are significantly higher or lower than expected by chance. If class assignment is consistent across the assessment configurations, the cell chi-square values on the diagonal would be higher than expected by chance, while the cell chi-square values on the off-diagonal would be lower than expected by chance.

Results revealed that only four participants were assigned to incongruent classes when comparing the 12 wave and 2-year interval configurations (Table 21).

When comparing the 12 wave configuration to either the configuration with uneven intervals (Table 22) or the first six waves (Table 23), approximately 14% of the sample was misclassified. For instance, 87 participants were assigned to Class 4 (High-Slowly Decreasing Class) when using all 12 waves, and these same participants were assigned to Class 3 (Low-Steady Class) when using uneven intervals.

Misclassification was almost always to the adjacent class that had lower drinking at Wave 1. When comparing the 12 wave and last seven wave configurations, 62% of the sample was misclassified (Table 24). Participants were assigned to adjacent and non-adjacent classes, such as Class 1 (Normative) using all 12 waves and Class 4 (High-Slowly Decreasing Class).

The contingency tables provide a key finding on the degree to which the participants were assigned to a similar class across the assessment configurations. Therefore, these results were compiled into Table 25 to underscore the degree of misclassification across the assessment configurations. Using the first row of the table as an example, 99.95% of the participants classified into Class 1 using 12 waves were also classified into Class 1 using 2-year intervals, while 0.05% of the participants classified into Class 1 using 12 waves were classified into Class 4 using 2-year intervals (Table 25). The 12 wave versus last seven wave comparison shows a substantial degree of misclassification across all five classes. For example, only 42.82% of the participants classified into Class 1 using 12 waves were also classified into Class 1 using the last seven waves. A portion of the results from Table 25 are graphically presented in Figure 11 to highlight the degree of misclassification in class assignment between the 12 wave versus last seven wave configurations. Within each

class from the 12 wave solution, participants were classified into any of the five classes using the last seven waves rather than being restricted to their corresponding class. This misclassification likely reflects substantial variation in the individual-level trajectories when trajectories were estimated over 12 waves versus the last seven waves.

Summary comparing assessment configurations. Six indicators of class agreement were used to compare the latent subgroups across the five assessment configurations: estimated class trajectories, class counts and proportions, average posterior probabilities, the estimated parameters from the five-class GMMs, Cohen's kappa, and contingency tables. Four key findings emerged from the applied study. First, all six indicators of class agreement suggested that the five-class GMM results were very similar when using all 12 waves versus 2-year intervals. Second, the five-class GMM results when comparing all 12 waves to either the configuration with uneven intervals or the first six waves showed some level of discrepancy. This discrepancy across the assessment configurations was most evident in the results for Cohen's kappa (Table 20) and the contingency tables (Table 25). Third, the five-class GMM results for the uneven intervals and first six wave configurations were very similar, which was particularly evident when evaluating Cohen's kappa (Table 20). Fourth, the largest discrepancy in the five-class GMM results was observed when comparing the 12 wave and last seven wave configurations. This discrepancy was evident across all six indicators used, particularly Cohen's kappa (Table 20) and the contingency table (Table 25).

Re-fitting the GMM. The model-building process that was used to determine the best fitting GMM using all 12 waves was conducted separately for each of the alternative assessment configurations. As detailed for the 12 wave configuration, the model-building process included fitting the latent growth models to determine the best fitting trajectory shape, then fitting the two-class through seven-class GMMs to determine the best fitting number of latent classes. By re-fitting the models, it can be determined whether the five-class GMM solution would have been considered the best fitting model for each of the four alternative configurations (Sher et al., 2011).

The series of latent growth models showed that the best fitting trajectory shape for each of the four alternative configurations included the following latent factors: intercept, linear slope, quadratic slope, and cubic slope. This result is the same as that determined using all 12 waves. The average estimated trajectory for the full sample was quite similar across all five assessment configurations (Figure 12). The best fitting GMM remained the five-class model for each of the four assessment configurations.

CHAPTER 4

SIMULATION STUDY: METHOD

Specific Aim 2

A Monte Carlo simulation study was conducted to evaluate how the timing and spacing of assessments affected the validity of the GMM results. The GMM results from the applied study were used to specify the population parameters in the data generation model using all 12 waves. The data analysis routines varied based on alterations in the timing and spacing of the assessments. To assess the degree to which the population parameters were recovered, the population parameters specified in the data generation model were compared to the parameter estimates obtained for each of the assessment configurations. Hypothesis 2a: GMM will exhibit variable performance at recovering the population parameters across the five assessment configurations. The configuration with all 12 waves will exhibit the best performance at recovering the population parameters with poorer performance for the four alternative configurations. Hypothesis 2b: The ability of GMM to recover the population parameters will diminish as sample size decreases.

Simulation Procedure

The following procedural steps were taken for the Monte Carlo simulation study (Fan et al., 2002; Mooney, 1997; Paxton et al., 2001). First, the empirical question was defined as determining how the timing and spacing of assessments affected the extent to which the population parameters of the GMM were recovered at various sample sizes. Defining a well-articulated research question is undoubtedly a critical step in the design of any simulation study since the research question guides

the analysis framework (Paxton et al., 2001). Second, the population parameters were defined on the basis of the results from the applied five-class GMM using 12 waves of alcohol data from the NLSY1997. Since the population characteristics were defined by the empirical results from the applied study, the external validity of the simulation study was enhanced (Paxton et al., 2001). Third, the two experimental factors for the simulation were defined as measurement timing and sample size, and the levels for each experimental condition were selected. Fourth, the data were computer-generated, for each sample size, according to the characteristics specified for the “true” population. Fifth, the data were analyzed using GMM for each of the experimental conditions. Sixth, the results were summarized in tables and figures for each experimental condition and for each parameter of interest.

Simulation Design

Experimental factors. The experimental factors used in the simulation were measurement timing (five levels) and total sample size (three levels), resulting in 15 experimental conditions. Recognizably, other factors may impact the accuracy of the alcohol use trajectory subgroups, such as the extent of missing data, variations in class proportions, or the degree of separation between the alcohol use trajectories of each latent class. When negotiating the tradeoff between the number of experimental factors and the number of levels for each factor, the advantages and disadvantages of various simulation designs must be evaluated in relation to the intended goal of the simulation study (Skrondal, 2000). The current research focused on measurement timing and sample size as an important initial investigation into the issue of measurement timing in GMM studies.

The data generation routine consisted of 500 replications whereby 500 samples were generated using 12 repeated-measurements, resulting in 1500 data sets (500 replications X 1 assessment configuration X 3 sample sizes). Data were only generated using a population with 12 repeated-measurements in order to evaluate the extent to which the population parameters (based on the 12 wave GMM solution from NLSY1997 data) were recovered across the assessment configurations. By using the same computer-generated data across the assessment configurations, any observed differences in parameter recovery may be attributed to alterations in the assessment configurations rather than to differences in the computer-generated data or the population model. A random seed number was specified in the data generation routine that enables the same data to be generated (e.g., for replication purposes).

The experimental conditions for measurement timing were the same five assessment configurations that were used in the applied study with the NLSY alcohol data (Table 2). By using the same assessment configurations for the applied study and the simulation study, the conclusions can be compared across the two studies.

Sample size was included as an experimental factor to determine the extent to which the validity of the GMM results was further impacted by differences in total sample size. The experimental conditions for sample size ($N = 500, 1000, 1500$) were selected in accordance with the range of sample sizes that were found in GMM or LCGA studies on alcohol. Among 31 applied alcohol studies on adolescents and young adults, sample size ranged from 81 to 32,087 participants (Figure 13). The median sample size was 489 and served as the basis for the minimum sample size in the simulation ($N = 500$). Lower sample sizes were not considered given that a five-

class GMM was being estimated and the smallest class, comprising 3.6% of the sample, would be approximately 18 participants with a sample size of 500. Sample sizes above the third quartile ($N = 2392$) were generally associated with larger, national studies. Therefore, based on the distribution of the sample sizes, $N = 1500$ was selected as the highest sample size for consideration in the current study. A sample size of 1000 was selected as the third and final level for sample size. This value represents the midpoint between the lowest ($N = 500$) and highest ($N = 1500$) sample sizes. As a result, there was the maximum degree of separation between the experimental levels, which would allow for more differentiated comparisons across the experimental conditions.

Population model for data generation. Data were generated for a five-class population. As in the applied study, the repeated-measurements (Y1 through Y12) were modeled based on the normal distribution, that is, reflecting continuous indicators. The values of the population parameters used in the data generation routine were based on the results of the applied study for the five-class GMM using all 12 waves. When the simulated data were first analyzed, the 12 wave configuration produced a high number of implausible values for the variance of the intercept factor (e.g., 131 of 500 replications had a negative variance for the intercept factor). The high proportion of implausible values in the replications may have been due to the very low population value for this parameter (i.e., 0.01). As a result the NLSY1997 data were re-analyzed with a five-class GMM in which the variance of the intercept factor and the cubic factor were fixed at zero. These factor variances were fixed in the applied study for certain assessment configurations (e.g., variance of the intercept

factor for the configuration with two-year intervals). Therefore, there was a basis for imposing these constraints in the population model for the simulation study.

Three 5-class populations were generated—one for each of the three experimental conditions for sample size (see sample syntax in Appendix A). All populations included 12 waves (Y1 through Y12). Population parameters included: 20 means for the latent factors (four per class: intercept, linear, quadratic, cubic), two factor variances (linear and quadratic), one covariance between the linear and quadratic factors), 12 residual variances for the observed outcomes (one for each of the 12 repeated-measurements, Y1 through Y12), and four means for the categorical latent factor to specify the class proportion for each of the five latent classes.

Overview of the Simulation Analyses

The simulated data sets for each of the 15 experimental conditions were analyzed by specifying a five-class GMM. The population values for the parameters were used as starting values in the data analysis models. To test Specific Aim 2, the criteria used to evaluate the performance of the GMMs were: 1) information on the number of models that failed to converge or had improper solutions (e.g., negative variances), 2) coverage estimates, which reflect the proportion of replications for which the true parameter value is contained in the 95% confidence interval for the estimated parameter, 3) the average of the estimated parameters across the replications, and 4) the average standard errors of each parameter across the replications. These criteria indicate the level of precision in parameter recovery across the assessment configurations.

Data were generated and analyzed using Mplus Version 6 (Muthén & Muthén, 1998-2010). Three data generation input files (one for each sample size) were written to generate the data with the 12 repeated measures (Y1-Y12) (see sample syntax in Appendix A). For the 12 wave assessment configuration, data were generated and analyzed in the same input file. For the four alternative assessment configurations, the simulated data were analyzed using a separate input file, so that the alternative assessment configurations could be implemented in the analysis model (see sample syntax in Appendix B).

CHAPTER 5

SIMULATION STUDY: RESULTS

Model Convergence

Data were generated from the population parameters derived from the five-class GMM conducted with the NLSY1997 alcohol data. Profile plots of the simulated data were examined to determine whether the simulated data resembled the applied data. Figure 14 was created from Class 1 of the first replication for the condition where $N = 500$. Figure 15 was created from Class 5 of the first replication for the condition where $N = 1500$. The profile plots of the simulated data depicted a clustering of the outcome values at Wave 1 and variability in the individual-level trajectories over the 12 waves. Therefore, the profile plots were similar to those generated from the NLSY1997 data. The simulated data did, however, include negative values since the data were generated from a normal distribution.

Estimation of the five-class GMMs was successful for all 7500 replications, combined across the 15 experimental conditions (500 replications X 5 assessment configurations X 3 sample sizes). No warnings were obtained (e.g., implausible parameter values, automatically fixed parameters) for any of the replications in three of the assessment configurations: 12 waves, 2-year intervals, or uneven intervals. However, warnings were obtained for replications using the first six waves and the last seven waves. In the experimental conditions analyzing only the first six waves, a negative value was obtained for the variance of the quadratic factor in 96 replications for $N = 1500$, 121 replications for $N = 1000$, and 152 replications for $N = 500$. In the experimental conditions analyzing only the last seven waves, a negative value was

obtained for the variance of the quadratic factor in 11 replications for $N = 1500$, 32 replications for $N = 1000$, and 84 replications for $N = 500$. Given that the quadratic factor was not fixed in the applied analyses, it was not justified to re-calculate the population parameters with this adjustment, which would have resulted in a free variance parameter for only the linear factor. Furthermore, adjustments were not made to the data analysis model (e.g., fixing the variance of the quadratic factor at zero), since this would have introduced an undesirable mismatch between the population model generating the data and the analysis model used to analyze the data. Consequently, any differences across the experimental conditions could be a result of this mismatch rather than the experimental factors under investigation. The data analysis models were re-estimated with a constraint that the variance of the quadratic factor be greater than zero; however, the warnings regarding the quadratic factor were not eliminated. Therefore, the original results for the first six waves and last seven waves were retained for further analysis and are reported here.

Parameter Recovery

The degree to which the population parameters were recovered in each of the 15 experimental conditions are detailed in a series of tables. A summary of the findings from these tables is presented here. Population parameters included: the four factor means for each of the five classes, the variances of the linear factor and the quadratic factor, the covariance between the linear and quadratic factors, and the 12 residual variances for each of the observed variables (Y1 through Y12). Tables 26 through 32 depict the results when $N = 1500$. Tables 33 through 39 depict the results when $N = 1000$. Tables 40 through 46 depict the results when $N = 500$. For each

parameter the following indices were evaluated: the population value for each parameter, the average of the estimated parameters, the average of the estimated standard errors, and 95% coverage for each parameter. First, the results pertaining to the 95% coverage estimates will be described for each set of parameters (i.e., class-specific means of the latent factors, variances and covariance for the latent factors, residual variances of the observed indicators). Then the results on the average of the estimated parameters and the average of the estimated standard errors will be described simultaneously for each set of parameters.

Results for 95% coverage. Coverage estimates reflect the proportion of replications (i.e., random samples) for which the true parameter value is contained in the 95% confidence interval for the estimated parameter. The degree to which the population parameter estimates were recovered across the replications is an important criterion for evaluating simulation studies.

Across four of the five assessment configurations, 95% coverage for the class-specific means of the latent factors (intercept, linear, quadratic, and cubic) was greater than .90, reflecting limited variability in the coverage estimates across the five classes. For the last seven wave configuration, 95% coverage estimates of the factor means were quite variable, ranging from .14 to .97. There were generally lower coverage estimates for the means of the intercept and linear factors and higher estimates for the means of the cubic factor. Coverage estimates were generally lower for the experimental conditions when sample size was equal to 500 compared to sample sizes equal to 1000 or 1500.

For the two variances and one covariance of the latent factors, four of the five assessment configurations had 95% coverage estimates ranging from .92 to .95 with limited variability across the levels of sample size. For the last seven waves only, 95% coverage estimates were below .52 for the variance of the intercept factor. In addition, the variance of the quadratic factor had 95% coverage estimates between .90 and .92, while the covariance had 95% coverage estimates between .85 and .88.

The vast majority of the 95% coverage estimates (across all five assessment configurations) for the residual variances ranged between .92 and .97 with only three values falling between .85 and .89.

Averages for the estimated parameters and standard errors. For the experimental factor of sample size ($N = 500, 1000, 1500$), the average of the estimated standard errors increased as the sample size decreased, and this held true across all of the estimated parameters. For the experimental factor of measurement timing, the results across the five assessment configurations are described separately for each estimated parameter.

The average of the estimated intercept factor means approximated the population values for all experimental conditions and classes, except for Classes 1 and 2 using only the last seven waves. Furthermore, the average of the estimated standard errors was high for all experimental conditions using the last seven waves. As an example, the average estimated means of the intercept factor for Class 1 are depicted in Figure 16 for all 15 experimental conditions, showing that the last seven wave configuration had the least accurate estimates and the largest standard errors.

The average of the estimated linear factor means were very close to the population values for all experimental conditions and classes, except for the last seven wave configuration. For certain classes the average estimate was lower than the population value (e.g., Class 1), and for other classes the average estimate was higher than the population value (e.g., Class 5). The average of the estimated standard errors was high for both the first six wave and last seven wave configurations. As an example, the average estimated means of the linear factor for Class 1 are depicted in Figure 17 for all 15 experimental conditions, showing that, once again, the last seven wave configuration had the least accurate estimates and the largest standard errors:

The average estimated quadratic factor means (Figure 18) and the average estimated cubic factor means (Figure 19) showed the same pattern. Specifically, average parameter estimates were the least accurate (i.e., high or low) for the first six wave and last seven wave configurations. Average estimates of the standard errors were also largest for these two assessment configurations.

The average estimated variances for the linear factor (Figure 20) and quadratic factor (Figure 21) were the least accurate for the last seven wave configurations, and the average standard errors were largest for the first six wave and last seven wave configurations. The results for the covariance between the linear and quadratic factor followed the same pattern as that described here for the variance parameters. Finally, the results for the residual variances were minimally affected by alterations in the assessment configurations or by total sample size.

CHAPTER 6

DISCUSSION

Overview

The applied study and simulation study examined the effect of measurement timing in the context of growth mixture modeling (GMM). The applied study addressed the reliability of GMM at identifying latent classes and at assigning participants to similar latent classes across five variations in the timing and spacing of survey assessments (i.e., assessment configurations). The simulation study addressed the validity of GMM by examining the model's ability to recover the population parameters across five assessment configurations. Key findings were that: 1) individuals' class assignment was least consistent when comparing the 12 wave and last seven wave configurations, and 2) the ability to recover the parameters from the 12 wave GMM solution was poorest when estimating only the first six waves and the last seven waves. Therefore, both the applied study and the simulation study provided some support for the hypothesis that variations in measurement timing may impact the reliability and validity of GMM solutions.

Summary of Findings for the Applied Study

Prospective longitudinal data were used to examine the reliability of GMM across five assessment configurations, which varied the timing and spacing of the assessments. Reliability was evaluated by comparing the GMM results with respect to the shapes of the average class trajectories and the degree of misclassification across the assessment configurations. The applied study used 12 waves of alcohol data from the NLSY1997 and identified latent classes based on the number of drinks participants

reported consuming per month. Overall, the GMM solutions were reliable when comparing the 12 wave versus two-year interval configurations (i.e., similar average class trajectories and very little misclassification). The reliability of the GMM solutions was more variable for the configurations with uneven intervals, the first six waves, and the last seven waves. There was markedly poorer reliability for the configuration using the last seven waves, compared to all 12 waves, given that 62% of the participants were classified into different (i.e., incongruent) classes across these conditions. Specific implications regarding the findings observed in the applied study are discussed next.

The five-class GMM solution. The five-class GMM using all 12 waves provided the best fit to these data and included the following classes: Class 1 (*Normative Class*, 71.33%), Class 2 (*Low-Increasing Class*, 8.45%), Class 3 (*Low-Steady Class*, 8.97%), Class 4 (*High-Slowly Decreasing Class*, 7.67%), and Class 5 (*Extreme-Sharply Decreasing Class*, 3.57%). Aside from the inclusion of a fifth class, these five classes differed from the four prototypical classes (i.e., consistently low use, increasing use, decreasing use, consistently high use) that have been observed across studies and were the focus of the investigation by Sher and colleagues (2011). Specifically, the average class trajectories observed with the NLSY1997 data converged over the last half of the study period rather than maintained a larger degree of differentiation among the classes.

The average trajectories of the five latent classes identified here differed from those identified in two applied alcohol studies that span similar developmental periods, specifically ages 13 to 23 years ($N = 3889$, Tucker et al., 2003) and ages 16 to

25 years ($N = 760$, Windle et al., 2005). The findings of these two studies are not directly comparable to the current findings, given that these studies used two different measures assessing heavy episodic drinking as the outcome while drinks per month served as the outcome here. Although heavy episodic drinking was assessed in the NLSY1997 sample, it was not used as the outcome measure due to low endorsement of this behavior across the study period. Nevertheless, these studies still serve as a meaningful comparison given the similarity in developmental timing. One of the five classes identified by Tucker and colleagues differed markedly in shape from those identified here, specifically a steady increasing group (16% of the sample) that exhibited their highest level of drinking at the final wave. Windle and colleagues found that the classes differed primarily in level of drinking across the study period with drinking, generally, increasing to a peak around age 20 and then decreasing. Although a four-class solution (i.e., consistently low, consistently high, increasing, decreasing) has commonly been found, classes with other trajectory shapes have also been identified. As noted by Sher and colleagues (2011), it remains unclear as to which substantive, methodological, and analytic factors may contribute to the identification of the common four-class solution. Presumably, the reverse is also the case in that it remains unclear as to which factors may contribute to the identification of latent classes with average trajectories that exhibit different types of patterns of change over time.

Results reported here suggested that the outcome at Wave 1 was especially influential in the estimation of the latent classes, given that the five classes were partitioned based on the observed values of the Wave 1 outcome measure (Table 13).

Findings from other studies demonstrate that the Wave 1 measure is not always so influential in the estimation of GMMs. For instance, the chronic and developmentally limited classes for heavy episodic drinking, identified by Jackson and Sher (2006) in a sample of emerging adults, had average Wave 1 estimates that were very similar to each other. In addition, three of the four classes identified by Tucker and colleagues (2003) had similar Wave 1 values with divergence across the study period in the average class trajectories. This suggests that there may have been a characteristic or set of characteristics associated with the individual-level trajectories estimated in the current sample that were particularly influential in demarcating the latent classes (e.g., Wave 1 outcome, value on the intercept factor). Furthermore, the individual-level trajectories within a given class differed substantially from the average class trajectories (Figure 6). A closer examination of how the individual-level parameter estimates for the trajectories (e.g., intercept, linear slope) are used to estimate the latent classes in Mplus and other software programs is warranted.

Summarizing across assessment configurations. Several observations regarding the assessment configurations are warranted with respect to model estimation, the patterns of the average class trajectories, and consistency in individuals' class membership. The two through five-class GMMs using all 12 waves converged without the need to address inadmissible solutions (e.g., negative variance estimates). However, the GMMs using the four alternative configurations required model modifications (e.g., fixing the variance of the intercept factor at zero) beginning with either the two- or three-class GMMs. For the two-class solution using only the first six waves, the residual variance of the Wave 1 outcome measure was very close

to zero (.004) and the R squared value for the Wave 1 measure was equal to .99 (standard error = .57). Therefore, the need to fix the residual variance in the three-class solution may have stemmed from the fact that class membership explained the variance associated with the Wave 1 indicator. Similarly, the variances of the intercept factor for the two-year interval and last seven wave configurations were not significantly different from zero in the two-class models, suggesting that the within-class variance was small after having taken class membership into account. Finally, for the two-class solution using uneven intervals, a correlation close to one was observed between the cubic and quadratic factors, which may have been a consequence of removing the latter waves, and thus, impacting the proper estimation of the cubic slope factor. Overall, variations in measurement timing (i.e., using fewer than 12 waves) required adjustments to the model being estimated such as fixing the variance of the intercept factor to zero. This suggests that having the additional waves of data prevented errors in model estimation given that the 2-class through 5-class GMMs using 12 waves estimated without the need to impose such constraints on the model (e.g., fix variance to zero).

Regarding class trajectories, in all five-class GMM solutions, the five latent classes had average trajectories that exhibited divergence at the first wave and relative convergence by the end of the study period (Figure 5). This general pattern, similar to a reverse “fanning effect,” was strikingly consistent across all five assessment configurations. Other applied studies that have altered the assessment configurations have also found similar average class trajectories despite the alterations to the waves included in the analyses (Jackson & Sher, 2006; Sher et al., 2011).

Differences did emerge, however, with respect to class assignment for the comparisons of the four alternative assessment configurations to the 12 wave configuration, which is also consistent with Jackson and Sher (2006). First, almost no misclassification occurred when comparing most likely class membership between the 12 wave and 2-year interval configurations. Second, when comparing the 12 wave configuration to either the uneven interval or first six wave configurations, 14% of the sample (approximately 360 participants) was misclassified. Third, the largest discrepancy in the five-class GMM results was observed when comparing the 12 wave and last seven wave configurations, resulting in misclassification of 62% of the sample. Given that participants were misclassified into adjacent (i.e., more similar) and non-adjacent (i.e., less similar) classes, the individual-level trajectories can vary substantially when the developmental window is more or less constricted (Figure 11). In fact, participants in any given class from the 12 wave solution were re-distributed across the five classes in the solution using the last seven waves, reflecting substantial incongruity in class assignment for these two assessment configurations. This finding suggests that, as would be expected, the developmental window under investigation substantially impacts the individual-level trajectories.

Across the 12 waves, participants' average ages ranged from 16 to 28 years. In contrast, for the last seven waves, participants' average ages ranged from 22 to 28 years, which is associated with the post-college years, career identification, and early adulthood (Brown et al., 2008). The high degree of misclassification suggests that individuals may exhibit markedly different drinking patterns at different developmental windows across their lifetime, especially during emerging adulthood

when heavy alcohol use may be incompatible with newly acquired social roles (e.g., marriage, parenthood) (Littlefield & Sher, 2010; Sher et al., 2005). On the individual-level, individuals exhibited markedly different patterns of alcohol use from adolescence to young adulthood. However, at the class-level, when GMMs using all 12 waves were estimated, the average drinking trajectories for each latent class were strikingly similar to each other during the period of young adulthood (Figure 5, Waves 7 to 12). Similarly, individual-level drinking patterns may also vary within a shorter timeframe (e.g., from semester to semester) suggesting that individuals' class assignment, for instance with a high school sample, may differ depending on whether the GMM includes a six month period or the entire year. This finding corroborates the observation from Sher and colleagues (2011) that there is a time-bounded component to the mixture modeling solutions. As observed by Jackson and Sher (2006), misclassification of participants across the assessment configurations may have been related, at least in part, to the probability of class membership and missing data.

These findings have implications regarding cost and efficiency for longitudinal research designs. The five-class GMM results were strikingly similar when using all 12 waves versus 2-year intervals, with respect to both the estimation of the average trajectories for each latent class and individuals' class assignments. The 2-year interval configuration included half as many assessments as the 12 wave configuration. This finding suggests that it would be cost effective to survey participants every two years instead of every year. In addition, the five-class GMM results, when comparing the uneven interval and first six wave configurations, were very similar. This finding suggests that the inclusion of Waves 8 and 12 for the uneven

interval configuration had a minimal effect on the estimation routine of the GMMs. Instead, the inclusion of Waves 1 through 4 in the uneven interval configuration approximated the results from Waves 1 through 6. This result suggests that the additional resources used for the longer term follow-ups may not have been necessary. The extent to which these findings generalize to other data sets and to the patterns observed in other phenomena over time has yet to be investigated.

Summary of Findings for the Simulation Study

A Monte Carlo simulation study was conducted to examine the validity of GMM across five assessment configurations, which varied the timing and spacing of the assessments. Validity was evaluated by comparing the GMM results on the degree to which the population parameters were recovered. Fifteen experimental conditions were tested (5 assessment configurations X 3 sample sizes). Overall, the validity in the GMM solution was supported for the following configurations: 12 waves, two-year intervals, and uneven intervals (i.e., high 95% coverage estimates, average estimated parameters that were close to the population values, and low average standard error estimates). Recall that the 95% coverage estimates indicate the proportion of replications for which the true parameter value is contained in the 95% confidence interval for the estimated parameter. The 95% coverage estimates were generally high for the configurations using the first six waves and last seven waves. However, the average estimated parameters were less accurate for the last seven wave configuration, and both of these conditions had high average standard error estimates. Specific comments regarding the findings observed in the simulation study are discussed next.

The simulation study showed that the 95% coverage estimates for the factor means, factor variances, and factor covariance were greater than .90 for four of the five assessment configurations, with the exception being the configuration using the last seven waves. The 95% coverage estimates of these parameters when using only the last seven waves ranged from .14 to .97, suggesting substantially less precision in the parameter estimates. In contrast, the 95% coverage estimates for the residual variances across all five assessment configurations were .92 or above for all but three of the 111 residual variances. Therefore, the coverage estimates of the residual variances were relatively unaffected by alterations in the assessment configurations.

Poor parameter recovery was observed for the assessment configurations using the first six waves and the last seven waves, but relatively good parameter recovery was observed using 12 waves, two-year intervals, and uneven intervals. This conclusion was based on the average for the estimated parameters (i.e., factor means, factor variances, and factor covariance) and the average for the standard errors. Generally, when using only the first six waves, the averages for the estimated parameters across replications were relatively close to the population values; however, the averages for the standard errors were large. Therefore, although the 95% coverage estimates using the first six waves were adequate and the parameter estimates were close to the population values, the large standard errors indicated a lack of precision (Figure 18). When using only the last seven waves, the averages for the estimated parameters were generally less accurate, and the averages for the standard errors were large. These findings provide some support for the hypothesis that variations in measurement timing may impact the validity of GMM solutions. In addition, the

precision of the variance and covariance parameter estimates decreased as overall sample size decreased, given the larger standard errors when sample size equaled 500 versus 1500. Similarly, the standard errors for the class-specific factor means increased as class size decreased (e.g., Class 5, the smallest class, had the largest standard errors).

Integrating the Applied and Simulation Studies

The two studies reported here examined the effect of measurement timing in the context of GMM. The applied study addressed the reliability of GMM at identifying latent classes and at assigning participants to similar latent classes across five assessment configurations. The key finding from the applied study was that the largest discrepancy in the five-class GMM results was observed when comparing the 12 wave and last seven wave configurations. The simulation study addressed the validity of GMM by examining the model's ability to recover the population parameters across five assessment configurations. A key finding from the simulation study was that the poorest parameter recovery was observed for the assessment configurations using the first six waves and the last seven waves. This finding is not surprising given that these two configurations produced substantially different parameter estimates in the applied study for the class-specific means of the slope factors, the slope factor variances, and the covariances among the factors (Table 19). Accordingly, it would be expected that the ability of these two assessment configurations to reproduce the parameter estimates associated with the five-class GMM using all 12 waves would be reduced. This incongruence may have also

contributed to the high rate of negative variance estimates for the quadratic factor in the replications under the first six wave and last seven wave configurations.

Strengths and Limitations

Strengths of the applied and simulation studies. The importance of properly spaced assessments to adequately capture changes in a given phenomena over time has been well recognized (Collins & Graham, 2002; Widaman, 1991; Wohlwill, 1973). The applied study reported here, which examines patterns of alcohol use from adolescence into young adulthood, contributes to the small body of research focusing on how measurement timing affects the reliability of the class solutions obtained from GMM (Eggleston et al., 2004; Jackson & Sher, 2006; Sher et al., 2011). Similarly, a limited number of simulation studies have examined the effect of the number of repeated-measurements in the context of mixture modeling analyses (Brown, 2003; Tolvanen, 2007). The simulation study reported here focused on the effect of alterations in interval spacing and timing on the validity of GMM. By consistently drawing from a population that was generated using all 12 waves, the simulation study was uniquely suited to examine the degree to which alterations in measurement timing would affect the technique's ability to recover the population parameters. Therefore, given the limited empirical literature that has directly examined variations in measurement timing in the context of GMM (applied studies or simulation studies), the current research provides a much-needed examination of this issue given the importance of measurement timing in longitudinal studies.

The NLSY1997 data set used was optimally suited to address the empirical question given that participants were assessed on 12 occasions over an 11.5 year

period. The inclusion of 12 assessments allowed for flexibility in the timing and spacing of the waves. The subsample analyzed here, which included 2686 participants, was sufficient for GMM. In fact, the sample size in the current study was relatively large compared to other applied studies implementing the same technique (Figure 13).

The applied study and simulation study are mutually informative, given that both studies utilized the NLSY alcohol data. Specifically, the population parameters used in the simulation study were derived from the results obtained in the applied study, thus increasing the external validity of the simulation study (Paxton et al., 2001). In addition, the same assessment configurations were tested in both studies to allow for direct comparisons of the findings between the two studies.

Limitations of the applied and simulation studies. The current study focused on the effect of measurement timing in the estimation of growth mixture models. Recognizably, other statistical techniques have been used to identify homogeneous subgroups of individuals with different patterns of behavior over time, including alcohol use. Latent class analysis for repeated-measures can be used to model binary outcome measures (e.g., no heavy episodic drinking versus any heavy episodic drinking) to ascertain patterns of heavy drinking over time (Lanza & Collins, 2006). Dynamic cluster analysis and times series analyses have also been used in combination to identify homogenous subgroups of individuals using smoking data collected daily for six weeks (Hoeppner, Goodwin, Velicer, Mooney, & Hatsukami, 2008). Lastly, cluster analysis has been used with ecological momentary assessment data to examine smoking patterns among non-daily smokers (Shiffman, Kirchner, Ferguson, & Scharf, 2009).

The outcome here was modeled as a continuous measure in both the applied (i.e., drinks per month) and simulation studies. However, other distributions may be suitable for these data, including modeling the outcome as count data with a zero-inflated Poisson or negative binomial distribution (Reinecke & Seddig, 2011). In preliminary analyses, these data were modeled using a Poisson and zero-inflated Poisson model in the estimation of the latent growth models. However, large values on the outcome produced difficulties in model estimation (e.g., automatically fixing the means, variances, and covariances of the latent factors), and subsequent analyses modeled the outcome as a continuous variable.

It has been noted that covariates can influence the estimation of the latent subgroups if the covariates are related to the growth parameters (intercept or slope) or the latent subgroups (Jung & Wickrama, 2008). However, covariates were not included in either the applied study or the simulation. Potential covariates that may be relevant in the estimation of latent subgroups with distinct patterns of alcohol use over time are discussed later in the section on “Future Research.”

Missing data were modeled in the applied study using maximum likelihood estimation and assuming data were missing at random (Muthén & Asparouhov, 2009; Muthén & Muthén, 1998-2010). However, the simulation study included complete cases without the consideration of missing data. Missingness was not imposed in the simulation study in order to first examine measurement timing without the additional noise introduced by imposing various levels of missingness. Therefore, the simulation findings may be considered a “best-case scenario” since complete cases were used.

Despite these limitations, the current findings are a useful contribution to the small body of literature examining measurement timing and the reliability and validity of GMM (Brown, 2003; Eggleston et al., 2004; Jackson & Sher, 2006, Sher et al., 2011; Tolvanen, 2007).

Future Research

The current findings suggest several fruitful avenues of future research. First, future applied work should examine the effect of measurement timing using different large-scale and small-scale data sets. Doing so would determine whether or not the current findings generalize to: 1) data sets with different patterns of change over time (i.e., different shapes for the average trajectories of each latent class), 2) samples with different characteristics (e.g., average age at Wave 1), 3) studies with different research designs (e.g., time lag between assessments), and 4) outcomes measuring different substance use behaviors (e.g., heavy episodic drinking, marijuana use).

Second, future applied work should determine which factors are associated with misclassification across different assessment configurations. Jackson and Sher (2006) found participants' probability of class membership and amount of missing data were associated with misclassification. Other relevant predictors of misclassification may also be identified. For instance, individuals who drank heavily, but infrequently, or individuals who reported greater fluctuations in their drinking over time may have been more likely to be misclassified in the 12 wave versus first six wave configurations. Similarly, individuals who were classified into different classes for the 12 wave versus last seven wave configurations may have exhibited markedly different patterns of alcohol use during these two timeframes. It may be useful to

apply piecewise GMMs in order to examine this type of misclassification, since piecewise GMMs would allow for different trajectory shapes, for example, over the first six and last seven waves (Li, Duncan, & Hops, 2001). A piecewise GMM using all 12 waves could then be compared to the first six and last seven wave configurations to determine whether the piecewise GMM reduces the proportion of participants who are misclassified across the assessment configurations.

Third, future applied and simulation studies should examine the combined effect that covariates and measurement timing may have on the reliability and validity of GMM. Relevant covariates that have been associated with alcohol use or alcohol use patterns include college attendance (Lanza & Collins, 2006; White et al., 2006), personality factors (Littlefield et al., 2009), peer influences (Borsari & Carey, 2001), parental monitoring and knowledge (Fairlie, Wood, & Laird, 2012; Patock-Peckham & Morgan-Lopez, 2010), and genetic factors (Zucker, 2006). An important consideration when adding covariates to predict class membership is whether to model the covariate as a time-invariant or time-varying variable (Bollen & Curran, 2006). Time-invariant covariates are generally assessed at one time point (e.g., Wave 1) and may be viewed as stable over time. In contrast, time-varying covariates are viewed as dynamic variables that change over time and, therefore, must be assessed at multiple time points. In the majority of studies, the time-varying covariate is assessed using the same measurement timing as the outcome. However, the time lag associated with the causal process between the covariate and the outcome must be considered in addition to the pattern of change in the outcome (Gollob & Reichardt, 1991; Haviland & Nagin, 2005). Therefore, decisions regarding measurement timing are further

complicated with the addition of covariates in the context of mixture modeling analyses.

Fourth, future simulation studies should use population values derived from the results of applied studies that used different samples, research designs, and outcomes in order to determine the extent to which the findings on the validity of GMM generalize across different data sets, research designs, and outcomes. In addition, simulation studies could be conducted to determine whether or not the correct population model is empirically identified as the best fitting model when variations in measurement timing are considered. For example, when data are generated under a four-class population model, the four-class model could be tested against (incorrect) three-class and five-class models under different assessment configurations to determine whether the (correct) four-class GMM is found to exhibit the best model fit. This type of analysis would attest to the robustness of GMM to variations in measurement timing.

Fifth, future simulation studies should examine how the validity of GMM results would be affected by measurement timing in combination with different levels of within-class variance estimates for the latent factors (see Bauer & Curran, 2003a). The applied study found that the variance estimates of the latent factors were larger for the first six wave and last seven wave configurations, which corresponds to the configurations in the simulation study that had the largest average estimates of the standard errors. Therefore, when estimated in conjunction with variations in measurement timing, the precision of the parameter estimates may decrease as the magnitude of the variance estimates increases. In addition, simulation studies could

examine how fixing the within-class variance to zero, as in LCGA, may alter the results associated with measurement timing. Other experimental factors could also be considered. For instance, the number of slope factors included may limit variation in the shapes of the average trajectories for each class. Parameter recovery may be negatively impacted if the population is generated using linear, quadratic, and cubic slopes, but the analysis model includes only a linear and quadratic slope. Similarly, the reliability of the GMM solution in applied studies may be negatively impacted by alterations in the slope parameters being estimated (Sher et al., 2011).

Conclusions

Adolescence and young adulthood are developmental periods during which there is substantial heterogeneity in alcohol use. This heterogeneity is evident when comparing consumption levels across individuals as well as individuals' patterns of alcohol use over time. A more nuanced understanding of individuals' patterns of alcohol use during adolescence is of critical importance for refining preventive interventions in this population. Growth mixture modeling (GMM) is a statistical technique that may be used to identify latent subgroups of individuals who may exhibit distinct patterns of alcohol use over time. Decisions about how to space assessments is particularly challenging in the context of GMM, since the latent subgroups may exhibit different trends in alcohol use.

Findings suggest that the developmental window under investigation had the most substantial impact on the comparability of the five-class GMM solutions across the assessment configurations. Individuals' class assignment in the 12 wave configuration was least consistent with individuals' class assignment in the last seven

wave configuration. Notably, variations in the length of the timeframe under investigation may produce less reliable GMM results if individuals' behavior across the study period shifts over time (e.g., from adolescence to young adulthood). In addition, the similarity of solutions for the 12 wave and two-year interval configurations suggests that it may, at times, be more cost effective to survey participants every two years instead of every year, depending on the nature of the research question being addressed. Future research examining the issue of measurement timing in conjunction with specific statistical techniques may further shed light on the conditions under which annual survey assessments may prove to be less cost effective.

It is crucial that researchers increase the available knowledge regarding: 1) the conditions that may contribute to the reliability of the latent classes identified using GMM and related techniques, and 2) the conditions that may increase the validity of GMM and related techniques via simulation studies. The time-bounded nature of the mixture modeling solutions (i.e., developmental window) suggests that the latent subgroups should not be interpreted as representing subgroups that are present in the population. Instead, the identification of latent subgroups is sensitive to variations in research design, which include, but may not be limited to, measurement timing. Further examinations of the intersection between statistical methods and the developmental course of alcohol use would enhance knowledge on the strengths and limitations of applying such techniques to the psychological literature. In doing so, it may then be possible to conduct more fine-grained examinations of individuals'

patterns of alcohol use in the context of prevention and intervention programs, etiological predictors, and distal outcomes.

Table 1
Participants' ages across the 12 waves.

Wave	N	Mean	SD	Median	Minimum	Maximum
1	2686	15.94	0.57	15.92	15.00	16.92
2	2473	17.57	0.63	17.58	15.58	19.00
3	2397	18.54	0.64	18.50	16.58	20.08
4	2346	19.61	0.63	19.58	17.67	21.08
5	2269	20.61	0.63	20.58	18.58	22.00
6	2266	21.59	0.63	21.58	19.58	23.25
7	2230	22.55	0.63	22.50	20.50	24.25
8	2168	23.59	0.63	23.58	21.58	25.08
9	2088	24.55	0.63	24.50	22.58	26.08
10	2187	25.51	0.63	25.50	23.58	27.00
11	2139	26.47	0.63	26.42	24.58	27.92
12	2164	27.50	0.63	27.50	25.58	29.08

Table 2

Assessment configurations for the applied study and simulation study.

Assessment Configuration	Wave											
	1	2	3	4	5	6	7	8	9	10	11	12
1: All waves	X	X	X	X	X	X	X	X	X	X	X	X
2: Two-year intervals	X		X		X		X		X		X	
3: Uneven intervals	X	X	X	X				X				X
4: First six waves	X	X	X	X	X	X						
5: Last seven waves						X	X	X	X	X	X	X

Table 3
Follow-up rate by wave.

Wave	Follow-up rate	Number of participants interviewed	Number of non-interviewed participants
1	100.00%	2686	0
2	92.07%	2473	213
3	89.24%	2397	289
4	87.34%	2346	340
5	84.48%	2269	417
6	84.36%	2266	420
7	83.02%	2230	456
8	80.71%	2168	518
9	77.74%	2088	598
10	81.42%	2187	499
11	79.64%	2139	547
12	80.57%	2164	522

Note. Follow-up rate is calculated as the percent of respondents from the baseline assessment who completed each follow-up.

Table 4

Total number of waves completed for each participant.

Number of completed waves	<i>n</i>	Percent of sample
1	51	1.90%
2	38	1.41%
3	52	1.94%
4	77	2.87%
5	67	2.49%
6	78	2.90%
7	81	3.02%
8	106	3.95%
9	119	4.43%
10	171	6.37%
11	329	12.25%
12	1517	56.48%

Table 5

Descriptive information for drinks per month (original scale).

Wave	<i>N</i>	% "0"	Mean	<i>SD</i>	Median	Min.	Max.	Skew	Kurtosis
1	2660	71%	8.46	42.20	0	0	1110	14.15	283.89
2	2453	58%	14.87	51.39	0	0	950	8.71	107.78
3	2363	49%	20.47	58.44	1	0	1080	8.40	108.17
4	2316	43%	25.18	65.16	2	0	900	6.00	51.51
5	2245	38%	32.87	98.70	4	0	1335	7.53	73.14
6	2257	31%	29.78	76.42	6	0	1155	7.05	68.30
7	2213	31%	31.23	94.25	6	0	1455	8.91	104.45
8	2112	31%	25.00	61.57	6	0	1050	8.13	96.98
9	2010	31%	25.02	63.10	6	0	1005	8.28	94.44
10	2114	29%	24.89	70.50	6	0	1170	9.75	126.27
11	2056	30%	22.13	58.53	6	0	1050	10.03	137.44
12	2079	32%	22.87	63.79	6	0	1140	9.90	132.72

Note. The % "0" refers to the percentage of participants who reported consuming zero drinks per month.

Table 6

Correlation coefficients for drinks per month using the original scale (upper triangle) and the logged scale (lower triangle).

Wave	1	2	3	4	5	6	7	8	9	10	11	12
1	-	.35	.15	.19	.11	.12	.19	.08	.08	.06	.07	.03 ^a
2	.45	-	.21	.22	.19	.19	.22	.13	.14	.16	.18	.10
3	.37	.53	-	.41	.18	.25	.08	.15	.13	.13	.19	.12
4	.33	.46	.57	-	.37	.30	.21	.22	.21	.27	.19	.18
5	.29	.40	.50	.61	-	.32	.29	.24	.28	.35	.18	.24
6	.25	.35	.44	.50	.57	-	.23	.35	.23	.29	.22	.20
7	.22	.32	.38	.48	.55	.59	-	.40	.30	.33	.33	.19
8	.22	.29	.37	.46	.49	.55	.63	-	.43	.35	.37	.18
9	.18	.28	.35	.41	.47	.51	.56	.62	-	.30	.39	.18
10	.17	.26	.37	.42	.47	.53	.58	.62	.63	-	.29	.31
11	.19	.30	.36	.42	.46	.52	.54	.60	.62	.68	-	.42
12	.15	.24	.31	.37	.43	.43	.47	.49	.54	.59	.67	-

Note. The correlations for drinks per month in the original scale appear in the upper triangle, and the correlations for the logged scale appear in the lower triangle. Sample size for the correlation coefficients ranges from 1,823 to 2,431.

^a Correlation is not significant. All other correlations are significant at $p < .05$.

Table 7

Descriptive information for drinks per month (logged scale).

Wave	<i>N</i>	Mean	<i>SD</i>	Median	Minimum	Maximum	Skew	Kurtosis
1	2660	0.69	1.29	0.00	0	7.01	1.92	2.89
2	2453	1.11	1.56	0.00	0	6.86	1.19	0.24
3	2363	1.44	1.70	0.69	0	6.99	0.79	-0.68
4	2316	1.64	1.76	1.10	0	6.80	0.63	-0.85
5	2245	1.85	1.80	1.61	0	7.20	0.50	-0.86
6	2257	2.01	1.71	1.95	0	7.05	0.28	-0.99
7	2213	2.00	1.72	1.95	0	7.28	0.32	-0.88
8	2112	1.97	1.66	1.95	0	6.96	0.22	-1.09
9	2010	1.98	1.64	1.95	0	6.91	0.21	-1.03
10	2114	1.96	1.62	1.95	0	7.07	0.26	-0.91
11	2056	1.93	1.59	1.95	0	6.96	0.21	-1.01
12	2079	1.89	1.62	1.95	0	7.04	0.26	-0.99

Table 8

Model selection indices for the latent growth models using 12 waves.

Model fit index	Intercept only	Intercept and linear	Intercept, linear and quadratic	Intercept, linear, quadratic, and cubic
Chi-square	5330.26	2565.03	678.57	293.75
df	76	73	69	64
CFI	0.58	0.80	0.95	0.98
TLI	0.64	0.82	0.95	0.98
AIC	95231.16	92471.93	90593.47	90218.65
BIC	95313.70	92572.16	90717.28	90371.94
aBIC	95269.22	92518.15	90650.56	90289.33
RMSEA	0.16	0.11	0.06	0.04
SRMR	0.20	0.10	0.04	0.03
Significant factor means?	Yes	Yes, both	Yes, all three	Yes, all four
Significant factor variances?	Yes	Yes, both	Yes, all three	Yes, all four

Table 9
Estimated parameters for the best fitting latent growth model using 12 waves.

Latent factor	<i>b</i>	<i>SE</i>	95% CI for <i>b</i>	β
Estimated mean parameters				
Intercept	0.67***	0.03	[0.62, 0.72]	0.64
Linear	3.97***	0.18	[3.62, 4.33]	0.63
Quadratic	-3.70***	0.38	[-4.45, -2.95]	-0.30
Cubic	1.00***	0.22	[0.57, 1.43]	0.15
Estimated variance parameters				
Intercept	1.11***	0.07	[0.97, 1.25]	1.00
Linear	39.89***	3.07	[33.87, 45.9]	1.00
Quadratic	154.38***	11.87	[131.11, 177.65]	1.00
Cubic	47.36***	3.73	[40.04, 54.67]	1.00
Estimated covariance parameters				
Linear-Intercept	-1.35***	0.41	[-2.16, -0.55]	-.20
Quadratic-Intercept	0.52	0.73	[-0.92, 1.96]	.04
Quadratic-Linear	-71.60***	5.82	[-83, -60.2]	-.91
Cubic-Intercept	0.04	0.38	[-0.71, 0.79]	.01
Cubic-Linear	35.27***	3.10	[29.2, 41.34]	.81
Cubic-Quadratic	-83.40***	6.57	[-96.28, -70.52]	-.98

*** $p < .001$.

Table 10

Model selection indices for the GMMs using 12 waves.

Classes	AIC	BIC	aBIC	Entropy	VLMR LRT (p-value)	LMR adj. LRT (p-value)
1 class	90218.65	90371.94	90289.33	n/a	n/a	n/a
2 classes	88299.71	88482.48	88383.98	0.953	$p < .0001$	$p < .0001$
3 classes	87106.27	87318.52	87204.14	0.968	$p < .01$	$p < .01$
4 classes	86154.61	86396.34	86266.07	0.980	0.04	0.04
5 classes	85625.39	85896.60	85750.44	0.979	0.10	0.10
6 classes	84792.23	85087.02	84928.15	0.987	0.04	0.04
7 classes	84604.49	84905.17	84743.13	0.991	0.17	0.18

Note. In the 6- and 7-class GMMs, the residual variance of the Wave 1 measure of the outcome was fixed at zero. In addition, for the 7-class GMM, the variance of the cubic factor was fixed at zero. VLMR LRT = Vuong-Lo-Mendell-Rubin Likelihood Ratio Test. LMR adj. LRT = Lo-Mendell-Rubin adjusted Likelihood Ratio Test.

Table 11

Class counts and proportions based on the estimated GMM using 12 waves.

GMM	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
2 classes							
Class count	2226.28	459.72					
Proportion	0.83	0.17					
3 classes							
Class count	2095.34	384.18	206.48				
Proportion	0.78	0.14	0.08				
4 classes							
Class count	2025.19	315.39	237.56	107.86			
Proportion	0.75	0.12	0.09	0.04			
5 classes							
Class count	1919.07	220.70	243.07	206.31	96.85		
Proportion	0.71	0.08	0.09	0.08	0.04		
6 classes							
Class count	1911.25	224.40	220.32	188.20	108.77	33.06	
Proportion	0.71	0.08	0.08	0.07	0.04	0.01	
7 classes							
Class count	1910.51	190.02	181.18	143.33	139.91	92.58	28.48
Proportion	0.71	0.07	0.07	0.05	0.05	0.03	0.01

Table 12

Average posterior probabilities for most likely class membership (row) by latent class (column) for 4- and 5-class GMMs using 12 waves.

Most Likely Class	Latent Class				
	1	2	3	4	5
4-class GMM					
1	0.998	0.002	0.001	0.000	n/a
2	0.006	0.975	0.019	0.000	n/a
3	0.000	0.033	0.95	0.017	n/a
4	0.002	0.001	0.031	0.966	n/a
5-class GMM					
1	0.997	0.001	0.001	0.001	0.000
2	0.036	0.950	0.014	0.000	0.000
3	0.000	0.012	0.979	0.009	0.000
4	0.001	0.001	0.011	0.979	0.008
5	0.000	0.000	0.000	0.013	0.987

Table 13

Drinks per month at Wave 1 by most likely class from the 5-class GMM using 12 waves.

Latent class	N	N missing	Mean	SD	Median	Minimum	Maximum
Class 1	1892	24	0.00	0.00	0.00	0.00	0.00
Class 2	227	0	0.95	0.27	1.10	0.69	1.39
Class 3	241	0	2.03	0.32	1.95	1.61	2.57
Class 4	204	2	3.27	0.38	3.24	2.71	4.04
Class 5	96	0	4.79	0.63	4.62	4.11	7.01

Table 14

Median drinks per month and heavy episodic drinking by class for the 5-class GMM using 12 waves.

Alcohol measure	Wave											
	1	2	3	4	5	6	7	8	9	10	11	12
Class 1												
Drinks/month	0	0	0	0	2	4	4	4	6	4	4	4
Heavy drinking	0	0	0	0	0	0	0	0	0	0	0	0
Class 2												
Drinks/month	2	2	5	5	9	10	10	15	10	12	12	12
Heavy drinking	0	0	0	0	0	0	0	0	0	0	0	0
Class 3												
Drinks/month	6	10	15	15	15	21	16	15	12	12	13	10
Heavy drinking	1	0	1	1	1	1	1	1	0	0	1	0
Class 4												
Drinks/month	25	15	20	21	30	25	20	20	20	15	15	15
Heavy drinking	3	1	2	2	2	2	1	2	1	1	1	1
Class 5												
Drinks/month	100	38	28	29	37	30	30	25	14	18	21	16
Heavy drinking	7	2	2	3	3	2	2	2	1	1	1	1

Note. Heavy drinking was defined as the number of days participants consumed five or more drinks on the same occasion in the past 30 days.

Table 15

Demographic information by class for the 5-class GMM using 12 waves.

Demographics	Class 1 % (n)	Class 2 % (n)	Class 3 % (n)	Class 4 % (n)	Class 5 % (n)
Sex					
Male	50.05 (959)	44.49 (101)	48.96 (118)	57.77 (119)	63.54 (61)
Female	49.95 (957)	55.51 (126)	51.04 (123)	42.23 (87)	36.46 (35)
Race					
White	67.18 (1275)	78.76 (178)	81.17 (194)	89.22 (182)	87.50 (84)
Black	19.60 (372)	13.72 (31)	7.95 (19)	4.90 (10)	5.21 (5)
Amer. Indian	0.63 (12)	0.0 (0)	0.84 (2)	0.49 (1)	1.04 (1)
Asian	3.16 (60)	2.21 (5)	0.42 (1)	0.49 (1)	1.04 (1)
Other	9.43 (179)	5.31 (12)	9.62 (23)	4.90 (10)	5.21 (5)
Hispanic					
No	85.33 (1629)	88.55 (201)	83.33 (200)	91.22 (187)	86.46 (83)
Yes	14.67 (280)	11.45 (26)	16.67 (40)	8.78 (18)	13.54 (13)
Region					
Northeast	18.63 (357)	20.26 (46)	17.84 (43)	21.84 (45)	20.83 (20)
North Central	26.62 (510)	24.67 (56)	24.90 (60)	25.24 (52)	29.17 (28)
South	35.23 (675)	37.00 (84)	30.29 (73)	27.67 (57)	32.29 (31)
West	19.52 (374)	18.06 (41)	26.97 (65)	25.24 (52)	17.71 (17)
Age at Wave 1					
Mean	15.91	15.93	16.00	16.05	16.17
SD	0.57	0.56	0.55	0.56	0.59

Note. Class 1 (Normative, $n = 1916$), Class 2 (Low-Increasing, $n = 227$), Class 3 (Low-Steady, $n = 241$), Class 4 (High-Slowly Decreasing, $n = 206$), and Class 5 (Extreme-Sharply Decreasing, $n = 96$). Class membership was based on assignment to the participants' most likely class. Amer. Indian = American Indian.

Table 16

Estimated parameters for the best fitting 5-class GMM using 12 waves.

Latent factor	<i>b</i>	<i>SE</i>	95% CI for <i>b</i>	β
Class 1: Estimated mean parameters				
Intercept	0.00	0.00	[0.00, 0.01]	0.03
Linear	5.45***	0.19	[5.08, 5.83]	0.82
Quadratic	-5.40***	0.42	[-6.23, -4.57]	-0.41
Cubic	1.70***	0.25	[1.22, 2.18]	0.23
Class 2: Estimated mean parameters				
Intercept	0.96	0.03	[0.91, 1.01]	9.68
Linear	4.46***	0.62	[3.25, 5.67]	0.67
Quadratic	-4.67***	1.41	[-7.43, -1.91]	-0.35
Cubic	1.52***	0.81	[-0.08, 3.12]	0.21
Class 3: Estimated mean parameters				
Intercept	2.02***	0.03	[1.96, 2.09]	20.37
Linear	1.90**	0.65	[0.63, 3.17]	0.29
Quadratic	-1.56	1.47	[-4.44, 1.33]	-0.12
Cubic	0.04	0.87	[-1.66, 1.74]	0.01
Class 4: Estimated mean parameters				
Intercept	3.27***	0.05	[3.18, 3.36]	32.89
Linear	-2.45***	0.73	[-3.88, -1.03]	-0.37
Quadratic	3.68*	1.56	[0.63, 6.74]	0.28
Cubic	-1.97*	0.88	[-3.69, -0.24]	-0.27
Class 5: Estimated mean parameters				
Intercept	4.79***	0.09	[4.61, 4.97]	48.21
Linear	-9.91***	1.19	[-12.24, -7.59]	-1.49
Quadratic	14.65***	2.54	[9.67, 19.62]	1.10
Cubic	-6.95***	1.46	[-9.82, -4.08]	-0.95
Estimated variance parameters				
Intercept	0.01	0.02	[-0.03, 0.04]	1.00
Linear	44.40***	2.31	[39.87, 48.93]	1.00
Quadratic	177.44***	12.03	[153.87, 201.01]	1.00
Cubic	53.25***	4.11	[45.19, 61.31]	1.00
Estimated covariance parameters				
Linear-Intercept	-0.01	0.10	[-0.21, 0.19]	-0.02
Quadratic-Intercept	0.02	0.17	[-0.31, 0.34]	0.01
Quadratic-Linear	-83.01***	5.13	[-93.07, -72.96]	-0.94
Cubic-Intercept	-0.01	0.08	[-0.17, 0.15]	-0.01
Cubic-Linear	41.43***	2.90	[35.75, 47.11]	0.85
Cubic-Quadratic	-95.17***	6.97	[-108.83, -81.51]	-0.98

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

Table 17

Class counts and proportions based on the estimated 5-class GMMs for all five assessment configurations.

GMM	Class 1	Class 2	Class 3	Class 4	Class 5
12 waves					
Class count	1919.07	220.70	243.07	206.31	96.85
Proportion	0.71	0.08	0.09	0.08	0.04
2-yr intervals					
Class count	1917.24	220.91	242.96	205.93	97.96
Proportion	0.71	0.08	0.09	0.08	0.04
Uneven intervals					
Class count	2021.60	255.80	191.16	144.46	71.98
Proportion	0.75	0.10	0.07	0.05	0.03
First 6 waves					
Class count	2021.55	256.46	192.75	143.99	70.24
Proportion	0.75	0.10	0.07	0.05	0.03
Last 7 waves					
Class count	885.87	453.13	614.23	451.07	100.70
Proportion	0.35	0.18	0.25	0.18	0.04

Table 18

Average posterior probabilities for most likely class membership (row) by latent class (column) for the 5-class GMMs for all five assessment configurations.

Most Likely Class	Latent Class				
	1	2	3	4	5
12 waves					
1	0.997	0.001	0.001	0.001	0.000
2	0.036	0.950	0.014	0.000	0.000
3	0.000	0.012	0.979	0.009	0.000
4	0.001	0.001	0.011	0.979	0.008
5	0.000	0.000	0.000	0.013	0.987
2-yr intervals					
1	0.996	0.001	0.001	0.001	0.000
2	0.035	0.951	0.014	0.000	0.000
3	0.000	0.011	0.979	0.010	0.000
4	0.001	0.000	0.009	0.982	0.008
5	0.000	0.000	0.000	0.014	0.986
Uneven intervals					
1	0.997	0.002	0.001	0.000	0.000
2	0.000	0.976	0.024	0.000	0.000
3	0.000	0.018	0.953	0.029	0.000
4	0.001	0.001	0.033	0.954	0.010
5	0.000	0.000	0.000	0.025	0.975
First 6 waves					
1	0.997	0.002	0.001	0.001	0.000
2	0.000	0.977	0.023	0.000	0.000
3	0.000	0.020	0.959	0.021	0.000
4	0.001	0.001	0.037	0.952	0.011
5	0.000	0.000	0.000	0.027	0.973
Last 7 waves					
1	0.932	0.041	0.020	0.006	0.001
2	0.036	0.897	0.067	0.000	0.000
3	0.017	0.072	0.820	0.085	0.006
4	0.003	0.005	0.103	0.856	0.033
5	0.000	0.000	0.000	0.095	0.905

Table 19

Estimated parameters for the 5-class GMMs for the four alternative assessment configurations.

Latent variable	2-yrs <i>b</i> (<i>SE</i>)	Uneven <i>b</i> (<i>SE</i>)	First 6 <i>b</i> (<i>SE</i>)	Last 7 <i>b</i> (<i>SE</i>)
Class 1: Estimated mean parameters				
Intercept	0.00 (0.00)	0.04 (0.00)	0.04 (0.00)	0.09 (0.02)
Linear	5.29 (0.25)	5.40 (0.21)	5.58 (0.40)	8.30 (0.55)
Quadratic	-4.86 (0.62)	-5.59 (0.50)	-6.92 (2.03)	-23.35 (2.28)
Cubic	1.31 (0.38)	1.91 (0.30)	4.49 (2.53)	20.37 (2.49)
Class 2: Estimated mean parameters				
Intercept	0.96 (0.03)	1.51 (0.03)	1.51 (0.03)	1.65 (0.12)
Linear	5.11 (0.83)	2.99 (0.74)	4.06 (1.49)	-0.01 (1.10)
Quadratic	-6.60 (2.09)	-2.73 (1.80)	-9.14 (7.28)	-0.74 (4.30)
Cubic	2.78 (1.32)	0.56 (1.07)	9.75 (8.79)	1.57 (4.59)
Class 3: Estimated mean parameters				
Intercept	2.02 (0.03)	2.60 (0.07)	2.61 (0.07)	2.88 (0.21)
Linear	3.08 (0.86)	-0.86 (1.06)	-4.93 (1.82)	-3.71 (0.92)
Quadratic	-5.04 (2.20)	3.25 (2.44)	26.08 (8.29)	11.59 (3.37)
Cubic	2.41 (1.40)	-2.38 (1.41)	-30.39 (9.77)	-11.81 (3.64)
Class 4: Estimated mean parameters				
Intercept	3.26 (0.05)	3.65 (0.09)	3.67 (0.09)	4.08 (0.27)
Linear	-2.14 (0.87)	-4.60 (1.14)	-10.53 (2.11)	-7.95 (1.35)
Quadratic	3.05 (2.08)	7.51 (2.60)	37.73 (9.78)	21.63 (4.50)
Cubic	-1.64 (1.28)	-3.83 (1.52)	-38.15 (11.76)	-19.62 (4.41)
Class 5: Estimated mean parameters				
Intercept	4.78 (0.10)	5.00 (0.11)	5.02 (0.12)	5.54 (0.39)
Linear	-9.20 (1.48)	-14.82 (1.76)	-24.54 (3.74)	-22.71 (6.80)
Quadratic	13.60 (3.66)	23.83 (4.14)	78.83 (18.12)	69.01 (22.35)
Cubic	-6.64 (2.29)	-11.33 (2.37)	-77.09 (21.65)	-64.52 (21.56)
Estimated variance parameters				
Intercept	Fixed at 0	0.03 (0.01)	0.05 (0.00)	Fixed at 0
Linear	44.08 (4.37)	15.07 (0.83)	112.35 (15.31)	97.29 (9.65)
Quadratic	178.77 (28.69)	9.56 (0.67)	2032.25 (377.66)	1028.02 (180.06)
Cubic	56.39 (11.91)	Fixed at 0	2779.99 (576.72)	908.93 (233.06)
Estimated covariance parameters				
Lin.-Int.	N/A	0.05 (0.06)	-0.04 (0.13)	N/A
Quad.-Int.	N/A	-0.05 (0.05)	-0.07 (0.68)	N/A
Quad.-Lin.	-83.12 (10.93)	-11.49 (0.72)	-422.84 (74.40)	-302.94 (40.50)
Cub.-Int.	N/A	N/A	0.15 (0.83)	N/A
Cub.-Lin.	42.52 (6.70)	N/A	422.76 (88.00)	262.92 (43.15)
Cub.-Quad.	-98.20 (18.19)	N/A	-2303.89 (460.1)	-942.42 (200.56)

Note. Int. = intercept. Lin. = linear. Quad. = quadratic. Cub. = cubic.

Table 20

Cohen's kappa to compare class agreement across assessment configurations.

Assessment Configuration	Assessment Configuration				
	12 waves	2-yr intervals	Uneven intervals	First 6 waves	Last 7 waves
12 waves	-				
2-yr intervals	0.99	-			
Uneven intervals	0.69	0.69	-		
First 6 waves	0.69	0.69	0.99	-	
Last 7 waves	0.10	0.10	0.08	0.08	-

Note. Unweighted kappas are shown. Sample size ranges from 2505 to 2685.

Table 21

Contingency table showing class agreement between the 5-class GMMs using 12 waves and 2-year intervals.

12 waves	2-year intervals					Marginal
	Class 1	Class 2	Class 3	Class 4	Class 5	
Class 1						
Frequency	1914	0	0	1	0	1915
Cell Chi-Square	219.33	161.90	171.89	143.50	69.18	
Cell Percent	71.28	0	0	0.04	0	71.32
Class 2						
Frequency	0	227	0	0	0	227
Cell Chi-Square	161.99	2250.20	20.38	17.25	8.20	
Cell Percent	0	8.45	0	0	0	8.45
Class 3						
Frequency	0	0	241	0	0	241
Cell Chi-Square	171.98	20.38	2224.60	18.31	8.71	
Cell Percent	0	0	8.98	0	0	8.98
Class 4						
Frequency	2	0	0	203	1	206
Cell Chi-Square	143.03	17.42	18.49	2242.60	5.58	
Cell Percent	0.07	0	0	7.56	0.04	7.67
Class 5						
Frequency	0	0	0	0	96	96
Cell Chi-Square	68.51	8.12	8.62	7.29	2468.80	
Cell Percent	0	0	0	0	3.58	3.58
Marginal						
Frequency	1916	227	241	204	97	2685
Percent	71.36	8.45	8.98	7.60	3.61	100

Note. $\chi^2(16, N = 2685) = 10656.22, p < .0001$; Phi coefficient = 1.99; Cohen's kappa = .99. Cell chi-squares that are bolded indicate cell frequencies that are significantly higher than expected by chance ($p < .05$). Non-bolded cell chi-squares indicate frequencies that are significantly lower than expected by chance ($p < .05$).

Table 22

Contingency table showing class agreement between the 5-class GMMs using 12 waves and uneven intervals.

12 waves	Uneven intervals					Marginal
	Class 1	Class 2	Class 3	Class 4	Class 5	
Class 1						
Frequency	1915	0	0	0	0	1915
Cell Chi-Square	152.34	181.87	134.09	101.99	51.35	
Cell Percent	71.32	0	0	0	0	71.32
Class 2						
Frequency	112	115	0	0	0	227
Cell Chi-Square	20.57	405.00	15.89	12.09	6.09	
Cell Percent	4.17	4.28	0	0	0	8.45
Class 3						
Frequency	0	140	101	0	0	241
Cell Chi-Square	181.94	599.22	419.40	12.84	6.46	
Cell Percent	0	5.21	3.76	0	0	8.98
Class 4						
Frequency	0	0	87	119	0	206
Cell Chi-Square	155.52	19.56	365.18	1063.70	5.52	
Cell Percent	0	0	3.24	4.43	0	7.67
Class 5						
Frequency	0	0	0	24	72	96
Cell Chi-Square	72.47	9.12	6.72	69.77	1872.30	
Cell Percent	0	0	0	0.89	2.68	3.58
Marginal						
Frequency	2027	255	188	143	72	2685
Percent	75.49	9.50	7.00	5.33	2.68	100

Note. $\chi^2(16, N = 2685) = 5941.03, p < .0001$; Phi coefficient = 1.49; Cohen's kappa = .69. Cell chi-squares that are bolded indicate cell frequencies that are significantly higher than expected by chance ($p < .05$). Non-bolded cell chi-squares indicate frequencies that are significantly lower than expected by chance ($p < .05$).

Table 23

Contingency table showing class agreement between the 5-class GMMs using 12 waves and the first six waves.

12 waves	First 6 waves					Marginal
	Class 1	Class 2	Class 3	Class 4	Class 5	
Class 1						
Frequency	1915	0	0	0	0	1915
Cell Chi-Square	151.81	181.87	134.09	102.70	49.93	
Cell Percent	71.32	0	0	0	0	71.32
Class 2						
Frequency	112	115	0	0	0	227
Cell Chi-Square	20.62	405.00	15.89	12.17	5.92	
Cell Percent	4.17	4.28	0	0	0	8.45
Class 3						
Frequency	0	140	101	0	0	241
Cell Chi-Square	182.03	599.22	419.40	12.93	6.28	
Cell Percent	0	5.21	3.76	0	0	8.98
Class 4						
Frequency	1	0	87	118	0	206
Cell Chi-Square	153.60	19.56	365.18	1035.40	5.37	
Cell Percent	0.04	0	3.24	4.39	0	7.67
Class 5						
Frequency	0	0	0	26	70	96
Cell Chi-Square	72.51	9.12	6.72	84.45	1820.30	
Cell Percent	0	0	0	0.97	2.61	3.58
Marginal						
Frequency	2028	255	188	144	70	2685
Percent	75.53	9.50	7.00	5.36	2.61	100

Note. $\chi^2 (16, N = 2685) = 5872.04, p < .0001$; Phi coefficient = 1.48; Cohen's kappa = .69. Cell chi-squares that are bolded indicate cell frequencies that are significantly higher than expected by chance ($p < .05$). Non-bolded cell chi-squares indicate frequencies that are significantly lower than expected by chance ($p < .05$).

Table 24

Contingency table showing class agreement between the 5-class GMMs using 12 waves and the last seven waves.

		Last 7 waves					
12 waves		Class 1	Class 2	Class 3	Class 4	Class 5	Marginal
Class 1							
Frequency		769	306	424	255	42	1796
Cell Chi-Square		17.89	0.49	2.34	13.38	7.86	
Cell Percent		30.70	12.22	16.93	10.18	1.68	71.70
Class 2							
Frequency		50	46	64	43	8	211
Cell Chi-Square		9.80	3.81	1.99	0.76	0.02	
Cell Percent		2.00	1.84	2.55	1.72	0.32	8.42
Class 3							
Frequency		41	27	77	61	15	221
Cell Chi-Square		19.94	2.33	7.70	11.79	6.28	
Cell Percent		1.64	1.08	3.07	2.44	0.60	8.82
Class 4							
Frequency		39	22	50	60	16	187
Cell Chi-Square		12.88	2.42	0.13	21.25	12.82	
Cell Percent		1.56	0.88	2.00	2.40	0.64	7.47
Class 5							
Frequency		22	9	22	28	9	90
Cell Chi-Square		3.72	2.23	0.03	8.88	10.28	
Cell Percent		0.88	0.36	0.88	1.12	0.36	3.59
Marginal							
Frequency		921	410	637	447	90	2505
Percent		36.77	16.37	25.43	17.84	3.59	100

Note. $\chi^2(16, N = 2505) = 181.03, p < .0001$; Phi coefficient = 0.27; Cohen's kappa = .10. Cell chi-squares that are bolded indicate cell frequencies that are significantly higher than expected by chance ($p < .05$). Non-bolded cell chi-squares indicate frequencies that are significantly lower than expected by chance when chi-square is greater than 3.84 for one degree of freedom ($p < .05$).

Table 25

Percentage of participants assigned to Classes 1 to 5 in the alternative assessment configurations (column) based on class membership using 12 waves (row).

Comparison	Alternative Configuration				
	Class 1	Class 2	Class 3	Class 4	Class 5
12 waves	2-yr intervals				
Class 1	99.95	0	0	0.05	0
Class 2	0	100	0	0	0
Class 3	0	0	100	0	0
Class 4	0.97	0	0	98.54	0.49
Class 5	0	0	0	0	100
12 waves	Uneven intervals				
Class 1	100	0	0	0	0
Class 2	49.34	50.66	0	0	0
Class 3	0	58.09	41.91	0	0
Class 4	0	0	42.23	57.77	0
Class 5	0	0	0	25.00	75.00
12 waves	First 6 waves				
Class 1	100	0	0	0	0
Class 2	49.34	50.66	0	0	0
Class 3	0	58.09	41.91	0	0
Class 4	0.49	0	42.23	57.28	0
Class 5	0	0	0	27.08	72.92
12 waves	Last 7 waves				
Class 1	42.82	17.04	23.61	14.20	2.34
Class 2	23.70	21.80	30.33	20.38	3.79
Class 3	18.55	12.22	34.84	27.60	6.79
Class 4	20.86	11.76	26.74	32.09	8.56
Class 5	24.44	10.00	24.44	31.11	10.00

Note. Percentages are row percentages, such that the values in each row sum to 100%.

Table 26

Simulation results for the Class 1 factor means across the five assessment configurations, $N = 1500$.

Class 1 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.94	0.09
Linear (5.420)	5.41 (0.19)	0.20	0.04	0.96	1.00
Quadratic (-5.313)	-5.29 (0.40)	0.41	0.16	0.95	1.00
Cubic (1.643)	1.63 (0.24)	0.24	0.06	0.95	1.00
2-yr intervals					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.94	0.09
Linear (5.420)	5.41 (0.28)	0.29	0.08	0.96	1.00
Quadratic (-5.313)	-5.30 (0.69)	0.71	0.47	0.96	1.00
Cubic (1.643)	1.64 (0.43)	0.45	0.19	0.96	0.95
Uneven intervals					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.94	0.09
Linear (5.420)	5.41 (0.23)	0.24	0.05	0.95	1.00
Quadratic (-5.313)	-5.31 (0.56)	0.54	0.31	0.93	1.00
Cubic (1.643)	1.65 (0.34)	0.32	0.11	0.93	1.00
First 6 waves					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.94	0.09
Linear (5.420)	5.42 (0.57)	0.56	0.32	0.95	1.00
Quadratic (-5.313)	-5.39 (2.87)	2.81	8.24	0.95	0.50
Cubic (1.643)	1.77 (3.54)	3.46	12.49	0.95	0.08
Last 7 waves					
Intercept (0.003)	-1.03 (0.42)	0.39	1.24	0.17	0.83
Linear (5.420)	0.59 (2.94)	3.09	31.95	0.55	0.05
Quadratic (-5.313)	0.02 (11.15)	11.87	152.50	0.93	0.06
Cubic (1.643)	3.26 (11.86)	12.68	143.03	0.94	0.06

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 27

Simulation results for the Class 2 factor means across the five assessment configurations, $N = 1500$.

Class 2 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (0.963)	0.96 (0.02)	0.02	0.00	0.95	1.00
Linear (4.533)	4.54 (0.58)	0.60	0.34	0.96	1.00
Quadratic (-4.836)	-4.85 (1.21)	1.24	1.45	0.96	0.98
Cubic (1.603)	1.61 (0.72)	0.73	0.51	0.96	0.59
2-yr intervals					
Intercept (0.963)	0.96 (0.02)	0.02	0.00	0.95	1.00
Linear (4.533)	4.56 (0.84)	0.88	0.71	0.96	1.00
Quadratic (-4.836)	-4.90 (2.04)	2.15	4.15	0.96	0.64
Cubic (1.603)	1.64 (1.28)	1.35	1.64	0.96	0.20
Uneven intervals					
Intercept (0.963)	0.96 (0.02)	0.02	0.00	0.95	1.00
Linear (4.533)	4.57 (0.70)	0.71	0.49	0.95	1.00
Quadratic (-4.836)	-4.93 (1.57)	1.63	2.48	0.95	0.87
Cubic (1.603)	1.66 (0.93)	0.97	0.87	0.96	0.39
First 6 waves					
Intercept (0.963)	0.96 (0.02)	0.02	0.00	0.95	1.00
Linear (4.533)	4.65 (1.68)	1.68	2.82	0.95	0.79
Quadratic (-4.836)	-5.41 (8.46)	8.49	71.76	0.95	0.10
Cubic (1.603)	2.25 (10.37)	10.43	107.72	0.95	0.06
Last 7 waves					
Intercept (0.963)	0.56 (0.31)	0.28	0.26	0.58	0.59
Linear (4.533)	0.73 (1.23)	1.32	15.96	0.18	0.08
Quadratic (-4.836)	-1.11 (4.68)	5.10	35.69	0.88	0.04
Cubic (1.603)	1.94 (4.94)	5.47	24.51	0.97	0.05

Note. Average Estimated S.E. = average estimated standard error, M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 28

Simulation results for the Class 3 factor means across the five assessment configurations, $N = 1500$.

Class 3 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.95	1.00
Linear (1.863)	1.84 (0.55)	0.55	0.30	0.96	0.92
Quadratic (-1.491)	-1.42 (1.11)	1.15	1.24	0.95	0.23
Cubic (0.010)	-0.03 (0.64)	0.68	0.41	0.95	0.05
2-yr intervals					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.95	1.00
Linear (1.863)	1.85 (0.87)	0.82	0.75	0.94	0.61
Quadratic (-1.491)	-1.48 (2.11)	2.00	4.46	0.94	0.14
Cubic (0.010)	0.01 (1.31)	1.26	1.70	0.95	0.05
Uneven intervals					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.95	1.00
Linear (1.863)	1.86 (0.67)	0.66	0.45	0.94	0.82
Quadratic (-1.491)	-1.46 (1.52)	1.52	2.30	0.95	0.18
Cubic (0.010)	-0.02 (0.90)	0.91	0.81	0.95	0.05
First 6 waves					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.94	1.00
Linear (1.863)	1.94 (1.58)	1.56	2.50	0.94	0.25
Quadratic (-1.491)	-1.92 (7.96)	7.91	63.40	0.94	0.07
Cubic (0.010)	0.55 (9.74)	9.73	94.92	0.95	0.05
Last 7 waves					
Intercept (2.021)	1.93 (0.30)	0.27	0.10	0.75	0.98
Linear (1.863)	0.81 (1.01)	1.00	2.12	0.80	0.14
Quadratic (-1.491)	-1.92 (3.97)	3.88	15.93	0.96	0.08
Cubic (0.010)	0.68 (4.17)	4.12	17.81	0.96	0.04

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 29

Simulation results for the Class 4 factor means across the five assessment configurations, $N = 1500$.

Class 4 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.95	1.00
Linear (-2.191)	-2.21 (0.63)	0.60	0.40	0.94	0.95
Quadratic (3.049)	3.08 (1.29)	1.24	1.66	0.94	0.68
Cubic (-1.610)	-1.62 (0.74)	0.73	0.54	0.94	0.59
2-yr intervals					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.95	1.00
Linear (-2.191)	-2.21 (0.90)	0.88	0.80	0.95	0.69
Quadratic (3.049)	3.07 (2.13)	2.16	4.55	0.94	0.30
Cubic (-1.610)	-1.62 (1.32)	1.35	1.73	0.96	0.22
Uneven intervals					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.95	1.00
Linear (-2.191)	-2.20 (0.71)	0.71	0.51	0.95	0.88
Quadratic (3.049)	3.07 (1.66)	1.64	2.75	0.95	0.48
Cubic (-1.610)	-1.62 (1.00)	0.98	0.99	0.94	0.41
First 6 waves					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.95	1.00
Linear (-2.191)	-2.12 (1.71)	1.67	2.93	0.95	0.27
Quadratic (3.049)	2.61 (8.70)	8.43	75.68	0.94	0.06
Cubic (-1.610)	-1.05 (10.74)	10.37	115.47	0.94	0.06
Last 7 waves					
Intercept (3.262)	3.33 (0.32)	0.28	0.11	0.77	1.00
Linear (-2.191)	0.85 (1.22)	1.34	10.77	0.33	0.10
Quadratic (3.049)	-2.92 (4.70)	5.11	57.69	0.79	0.08
Cubic (-1.610)	-0.15 (4.97)	5.40	26.77	0.95	0.04

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 30

Simulation results for the Class 5 factor means across the five assessment configurations, $N = 1500$.

Class 5 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (4.785)	4.79 (0.03)	0.03	0.00	0.96	1.00
Linear (-9.614)	-9.64 (0.87)	0.88	0.76	0.94	1.00
Quadratic (13.969)	13.95 (1.82)	1.81	3.32	0.95	1.00
Cubic (-6.580)	-6.55 (1.08)	1.06	1.16	0.95	1.00
2-yr intervals					
Intercept (4.785)	4.78 (0.03)	0.03	0.00	0.95	1.00
Linear (-9.614)	-9.64 (1.35)	1.28	1.82	0.94	1.00
Quadratic (13.969)	13.97 (3.34)	3.14	11.15	0.93	0.98
Cubic (-6.580)	-6.57 (2.10)	1.97	4.40	0.93	0.90
Uneven intervals					
Intercept (4.785)	4.79 (0.03)	0.03	0.00	0.95	1.00
Linear (-9.614)	-9.61 (1.05)	1.04	1.10	0.93	1.00
Quadratic (13.969)	13.87 (2.45)	2.38	6.01	0.94	1.00
Cubic (-6.580)	-6.51 (1.48)	1.42	2.18	0.93	0.99
First 6 waves					
Intercept (4.785)	4.78 (0.03)	0.03	0.00	0.95	1.00
Linear (-9.614)	-9.47 (2.50)	2.44	6.28	0.93	0.96
Quadratic (13.969)	13.18 (12.56)	12.33	158.01	0.94	0.19
Cubic (-6.580)	-5.76 (15.29)	15.17	233.96	0.94	0.06
Last 7 waves					
Intercept (4.785)	4.95 (0.49)	0.40	0.27	0.85	1.00
Linear (-9.614)	0.78 (3.50)	3.44	120.29	0.14	0.06
Quadratic (13.969)	-3.64 (13.96)	13.21	504.75	0.61	0.05
Cubic (-6.580)	-1.54 (15.22)	14.03	256.77	0.94	0.04

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 31

Simulation results for the factor variance and covariance parameters across the five assessment configurations, $N = 1500$.

Factor Variances and Covariance (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Linear (13.572)	13.54 (0.65)	0.63	0.43	0.94	1.00
Quadratic (8.401)	8.38 (0.48)	0.46	0.23	0.94	1.00
Linear-Quadratic (-10.189)	-10.17 (0.55)	0.52	0.30	0.93	1.00
2-yr intervals					
Linear (13.572)	13.55 (0.88)	0.85	0.77	0.95	1.00
Quadratic (8.401)	8.39 (0.81)	0.78	0.65	0.93	1.00
Linear-Quadratic (-10.189)	-10.17 (0.81)	0.78	0.65	0.93	1.00
Uneven intervals					
Linear (13.572)	13.54 (0.87)	0.84	0.75	0.93	1.00
Quadratic (8.401)	8.38 (0.75)	0.71	0.57	0.93	1.00
Linear-Quadratic (-10.189)	-10.17 (0.76)	0.73	0.58	0.93	1.00
First 6 waves					
Linear (13.572)	13.49 (2.23)	2.18	4.98	0.94	1.00
Quadratic (8.401)	7.95 (9.19)	8.93	84.57	0.93	0.16
Linear-Quadratic (-10.189)	-10.00 (4.32)	4.18	18.62	0.93	0.65
Last 7 waves					
Linear (13.572)	6.06 (2.19)	2.64	61.23	0.15	0.69
Quadratic (8.401)	13.83 (6.48)	7.57	71.34	0.90	0.50
Linear-Quadratic (-10.189)	-6.83 (3.65)	4.35	24.53	0.87	0.36

Note. For the data generation model and the data analysis model, the variances of the intercept and cubic factors were fixed at zero. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 32

Simulation results for the residual variances of the observed indicators (Y1-Y12) across the five assessment configurations, N = 1500.

Residual Variances (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.95	1.00
Y2 (1.634)	1.64 (0.06)	0.06	0.00	0.94	1.00
Y3 (1.771)	1.77 (0.07)	0.07	0.00	0.96	1.00
Y4 (1.628)	1.63 (0.07)	0.06	0.00	0.94	1.00
Y5 (1.515)	1.51 (0.06)	0.06	0.00	0.94	1.00
Y6 (1.303)	1.30 (0.05)	0.05	0.00	0.94	1.00
Y7 (1.214)	1.22 (0.05)	0.05	0.00	0.95	1.00
Y8 (1.144)	1.14 (0.05)	0.05	0.00	0.96	1.00
Y9 (1.183)	1.18 (0.05)	0.05	0.00	0.96	1.00
Y10 (0.935)	0.93 (0.04)	0.04	0.00	0.95	1.00
Y11 (0.742)	0.74 (0.04)	0.04	0.00	0.96	1.00
Y12 (0.956)	0.96 (0.05)	0.05	0.00	0.95	1.00
2-yr intervals					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.94	1.00
Y3 (1.771)	1.77 (0.07)	0.07	0.01	0.96	1.00
Y5 (1.515)	1.51 (0.07)	0.07	0.00	0.94	1.00
Y7 (1.214)	1.22 (0.06)	0.06	0.00	0.94	1.00
Y9 (1.183)	1.18 (0.05)	0.06	0.00	0.97	1.00
Y11 (0.742)	0.74 (0.08)	0.08	0.01	0.96	1.00
Uneven intervals					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.95	1.00
Y2 (1.634)	1.64 (0.07)	0.06	0.00	0.93	1.00
Y3 (1.771)	1.78 (0.07)	0.07	0.01	0.95	1.00
Y4 (1.628)	1.63 (0.08)	0.07	0.01	0.94	1.00
Y8 (1.144)	1.13 (0.07)	0.07	0.00	0.94	1.00
Y12 (0.956)	0.95 (0.18)	0.17	0.03	0.94	1.00
First 6 waves					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.94	1.00
Y2 (1.634)	1.64 (0.07)	0.06	0.00	0.93	1.00
Y3 (1.771)	1.77 (0.07)	0.07	0.01	0.96	1.00
Y4 (1.628)	1.63 (0.07)	0.07	0.01	0.95	1.00
Y5 (1.515)	1.51 (0.07)	0.07	0.01	0.94	1.00
Y6 (1.303)	1.30 (0.10)	0.10	0.01	0.95	1.00
Last 7 waves					
Y6 (1.303)	1.23 (0.07)	0.08	0.01	0.85	1.00
Y7 (1.214)	1.23 (0.06)	0.06	0.00	0.95	1.00
Y8 (1.144)	1.15 (0.05)	0.05	0.00	0.96	1.00

Y9 (1.183)	1.18 (0.05)	0.05	0.00	0.96	1.00
Y10 (0.935)	0.92 (0.04)	0.04	0.00	0.93	1.00
Y11 (0.742)	0.74 (0.04)	0.04	0.00	0.95	1.00
Y12 (0.956)	0.94 (0.06)	0.07	0.00	0.95	1.00

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 33

Simulation results for the Class 1 factor means across the five assessment configurations, $N = 1000$.

Class 1 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.07
Linear (5.420)	5.41 (0.24)	0.24	0.06	0.95	1.00
Quadratic (-5.313)	-5.31 (0.48)	0.50	0.23	0.96	1.00
Cubic (1.643)	1.65 (0.28)	0.30	0.08	0.96	1.00
2-yr intervals					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.06
Linear (5.420)	5.41 (0.35)	0.36	0.12	0.95	1.00
Quadratic (-5.313)	-5.30 (0.86)	0.87	0.74	0.95	1.00
Cubic (1.643)	1.64 (0.54)	0.55	0.29	0.95	0.85
Uneven intervals					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.07
Linear (5.420)	5.42 (0.28)	0.29	0.08	0.95	1.00
Quadratic (-5.313)	-5.33 (0.64)	0.67	0.41	0.96	1.00
Cubic (1.643)	1.66 (0.38)	0.40	0.15	0.95	0.99
First 6 waves					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.07
Linear (5.420)	5.43 (0.70)	0.68	0.49	0.94	1.00
Quadratic (-5.313)	-5.41 (3.54)	3.44	12.48	0.95	0.35
Cubic (1.643)	1.79 (4.31)	4.23	18.52	0.95	0.07
Last 7 waves					
Intercept (0.003)	-1.02 (0.51)	0.41	1.30	0.25	0.75
Linear (5.420)	0.42 (3.60)	3.93	37.88	0.63	0.06
Quadratic (-5.313)	0.60 (14.21)	14.94	236.55	0.92	0.07
Cubic (1.643)	2.74 (15.40)	15.86	237.91	0.94	0.07

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 34

Simulation results for the Class 2 factor means across the five assessment configurations, $N = 1000$.

Class 2 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (0.963)	0.96 (0.03)	0.03	0.00	0.95	1.00
Linear (4.533)	4.54 (0.74)	0.73	0.54	0.94	1.00
Quadratic (-4.836)	-4.86 (1.49)	1.51	2.23	0.95	0.88
Cubic (1.603)	1.62 (0.87)	0.88	0.76	0.95	0.43
2-yr intervals					
Intercept (0.963)	0.96 (0.03)	0.03	0.00	0.94	1.00
Linear (4.533)	4.58 (1.12)	1.07	1.24	0.94	0.98
Quadratic (-4.836)	-4.96 (2.72)	2.62	7.37	0.94	0.49
Cubic (1.603)	1.68 (1.70)	1.64	2.89	0.93	0.18
Uneven intervals					
Intercept (0.963)	0.96 (0.03)	0.03	0.00	0.95	1.00
Linear (4.533)	4.56 (0.86)	0.87	0.73	0.94	1.00
Quadratic (-4.836)	-4.92 (1.90)	1.99	3.61	0.95	0.73
Cubic (1.603)	1.66 (1.12)	1.19	1.25	0.96	0.26
First 6 waves					
Intercept (0.963)	0.96 (0.03)	0.03	0.00	0.95	1.00
Linear (4.533)	4.67 (2.05)	2.06	4.22	0.95	0.60
Quadratic (-4.836)	-5.54 (10.36)	10.38	107.71	0.95	0.08
Cubic (1.603)	2.46 (12.67)	12.75	160.87	0.95	0.05
Last 7 waves					
Intercept (0.963)	0.59 (0.35)	0.28	0.26	0.59	0.58
Linear (4.533)	0.68 (1.49)	1.63	17.11	0.31	0.07
Quadratic (-4.836)	-0.93 (5.79)	6.34	48.68	0.92	0.03
Cubic (1.603)	1.73 (6.16)	6.72	37.90	0.97	0.04

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 35

Simulation results for the Class 3 factor means across the five assessment configurations, $N = 1000$.

Class 3 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.95	1.00
Linear (1.863)	1.84 (0.69)	0.68	0.48	0.95	0.78
Quadratic (-1.491)	-1.44 (1.42)	1.40	2.03	0.94	0.19
Cubic (0.010)	-0.02 (0.83)	0.82	0.69	0.95	0.06
2-yr intervals					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.95	1.00
Linear (1.863)	1.83 (1.01)	1.00	1.02	0.93	0.46
Quadratic (-1.491)	-1.45 (2.48)	2.45	6.12	0.94	0.10
Cubic (0.010)	-0.01 (1.54)	1.53	2.36	0.95	0.05
Uneven intervals					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.95	1.00
Linear (1.863)	1.86 (0.82)	0.80	0.67	0.93	0.65
Quadratic (-1.491)	-1.48 (1.90)	1.85	3.61	0.94	0.13
Cubic (0.010)	0.00 (1.15)	1.11	1.31	0.94	0.06
First 6 waves					
Intercept (2.021)	2.02 (0.02)	0.02	0.00	0.95	1.00
Linear (1.863)	1.97 (1.95)	1.91	3.80	0.93	0.20
Quadratic (-1.491)	-2.19 (9.84)	9.67	97.10	0.93	0.08
Cubic (0.010)	0.93 (12.09)	11.89	146.77	0.92	0.08
Last 7 waves					
Intercept (2.021)	1.96 (0.34)	0.28	0.12	0.73	0.97
Linear (1.863)	0.86 (1.63)	1.34	3.65	0.86	0.11
Quadratic (-1.491)	-2.15 (6.33)	5.25	40.47	0.94	0.08
Cubic (0.010)	0.88 (6.82)	5.60	47.20	0.94	0.06

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 36

Simulation results for the Class 4 factor means across the five assessment configurations, $N = 1000$.

Class 4 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.94	1.00
Linear (-2.191)	-2.19 (0.76)	0.73	0.58	0.92	0.84
Quadratic (3.049)	3.06 (1.53)	1.52	2.34	0.94	0.55
Cubic (-1.610)	-1.61 (0.88)	0.89	0.77	0.95	0.46
2-yr intervals					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.95	1.00
Linear (-2.191)	-2.21 (1.10)	1.07	1.22	0.95	0.55
Quadratic (3.049)	3.12 (2.68)	2.63	7.18	0.94	0.23
Cubic (-1.610)	-1.66 (1.68)	1.65	2.81	0.95	0.16
Uneven intervals					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.94	1.00
Linear (-2.191)	-2.19 (0.88)	0.87	0.77	0.94	0.70
Quadratic (3.049)	3.04 (2.03)	2.00	4.11	0.94	0.36
Cubic (-1.610)	-1.60 (1.22)	1.20	1.48	0.94	0.28
First 6 waves					
Intercept (3.262)	3.26 (0.02)	0.02	0.00	0.95	1.00
Linear (-2.191)	-2.10 (2.10)	2.05	4.39	0.94	0.20
Quadratic (3.049)	2.60 (10.73)	10.32	115.09	0.95	0.06
Cubic (-1.610)	-1.09 (13.30)	12.70	176.93	0.93	0.06
Last 7 waves					
Intercept (3.262)	3.35 (0.37)	0.28	0.15	0.74	1.00
Linear (-2.191)	0.84 (1.55)	1.72	11.56	0.51	0.08
Quadratic (3.049)	-2.94 (5.92)	6.61	70.84	0.85	0.06
Cubic (-1.610)	-0.10 (6.25)	7.01	41.25	0.96	0.04

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 37

Simulation results for the Class 5 factor means across the five assessment configurations, $N = 1000$.

Class 5 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (4.785)	4.79 (0.03)	0.03	0.00	0.95	1.00
Linear (-9.614)	-9.62 (1.08)	1.06	1.17	0.93	1.00
Quadratic (13.969)	13.92 (2.27)	2.21	5.15	0.94	1.00
Cubic (-6.580)	-6.54 (1.33)	1.29	1.78	0.94	1.00
2-yr intervals					
Intercept (4.785)	4.79 (0.03)	0.03	0.00	0.94	1.00
Linear (-9.614)	-9.59 (1.66)	1.56	2.74	0.93	1.00
Quadratic (13.969)	13.86 (4.03)	3.83	16.19	0.92	0.94
Cubic (-6.580)	-6.50 (2.51)	2.41	6.28	0.93	0.74
Uneven intervals					
Intercept (4.785)	4.79 (0.03)	0.03	0.00	0.94	1.00
Linear (-9.614)	-9.57 (1.30)	1.26	1.68	0.93	1.00
Quadratic (13.969)	13.78 (2.96)	2.90	8.78	0.94	1.00
Cubic (-6.580)	-6.46 (1.77)	1.73	3.15	0.93	0.95
First 6 waves					
Intercept (4.785)	4.79 (0.03)	0.03	0.00	0.94	1.00
Linear (-9.614)	-9.34 (3.11)	2.98	9.75	0.94	0.87
Quadratic (13.969)	12.64 (15.62)	15.09	245.22	0.95	0.15
Cubic (-6.580)	-5.20 (19.16)	18.57	368.09	0.95	0.07
Last 7 waves					
Intercept (4.785)	4.98 (0.58)	0.44	0.37	0.80	1.00
Linear (-9.614)	0.77 (4.56)	4.24	128.65	0.26	0.06
Quadratic (13.969)	-3.35 (17.78)	16.30	615.42	0.70	0.06
Cubic (-6.580)	-2.08 (19.17)	17.27	386.99	0.93	0.06

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 38

Simulation results for the factor variance and covariance parameters across the five assessment configurations, $N = 1000$.

Factor Variances and Covariance (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Linear (13.572)	13.52 (0.81)	0.77	0.66	0.92	1.00
Quadratic (8.401)	8.37 (0.59)	0.56	0.35	0.93	1.00
Linear-Quadratic (-10.189)	-10.15 (0.68)	0.64	0.46	0.93	1.00
2-yr intervals					
Linear (13.572)	13.56 (1.09)	1.04	1.18	0.94	1.00
Quadratic (8.401)	8.40 (0.99)	0.96	0.98	0.95	1.00
Linear-Quadratic (-10.189)	-10.19 (1.00)	0.96	0.99	0.95	1.00
Uneven intervals					
Linear (13.572)	13.53 (1.05)	1.03	1.10	0.93	1.00
Quadratic (8.401)	8.36 (0.92)	0.87	0.85	0.95	1.00
Linear-Quadratic (-10.189)	-10.15 (0.93)	0.89	0.86	0.93	1.00
First 6 waves					
Linear (13.572)	13.52 (2.70)	2.67	7.29	0.95	1.00
Quadratic (8.401)	8.22 (11.22)	10.93	125.70	0.93	0.11
Linear-Quadratic (-10.189)	-10.11 (5.25)	5.12	27.48	0.94	0.48
Last 7 waves					
Linear (13.572)	5.77 (2.79)	3.19	68.67	0.29	0.46
Quadratic (8.401)	13.11 (8.37)	9.19	92.13	0.92	0.31
Linear-Quadratic (-10.189)	-6.39 (4.69)	5.27	36.37	0.88	0.24

Note. For the data generation model and the data analysis model, the variances of the intercept and cubic factors were fixed at zero. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 39

Simulation results for the residual variances of the observed indicators (Y1-Y12) across the five assessment configurations, N = 1000.

Residual Variances (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.95	1.00
Y2 (1.634)	1.64 (0.08)	0.08	0.01	0.95	1.00
Y3 (1.771)	1.78 (0.08)	0.08	0.01	0.95	1.00
Y4 (1.628)	1.63 (0.08)	0.08	0.01	0.93	1.00
Y5 (1.515)	1.51 (0.08)	0.08	0.01	0.92	1.00
Y6 (1.303)	1.30 (0.06)	0.07	0.00	0.96	1.00
Y7 (1.214)	1.22 (0.06)	0.06	0.00	0.94	1.00
Y8 (1.144)	1.14 (0.06)	0.06	0.00	0.94	1.00
Y9 (1.183)	1.18 (0.06)	0.06	0.00	0.95	1.00
Y10 (0.935)	0.93 (0.05)	0.05	0.00	0.95	1.00
Y11 (0.742)	0.74 (0.04)	0.04	0.00	0.95	1.00
Y12 (0.956)	0.96 (0.07)	0.06	0.00	0.94	1.00
2-yr intervals					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.95	1.00
Y3 (1.771)	1.77 (0.09)	0.09	0.01	0.93	1.00
Y5 (1.515)	1.51 (0.08)	0.08	0.01	0.94	1.00
Y7 (1.214)	1.22 (0.08)	0.07	0.01	0.94	1.00
Y9 (1.183)	1.18 (0.07)	0.07	0.00	0.94	1.00
Y11 (0.742)	0.74 (0.10)	0.10	0.01	0.95	1.00
Uneven intervals					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.95	1.00
Y2 (1.634)	1.64 (0.08)	0.08	0.01	0.95	1.00
Y3 (1.771)	1.78 (0.09)	0.09	0.01	0.95	1.00
Y4 (1.628)	1.63 (0.09)	0.09	0.01	0.92	1.00
Y8 (1.144)	1.13 (0.08)	0.08	0.01	0.94	1.00
Y12 (0.956)	0.96 (0.22)	0.21	0.05	0.93	0.99
First 6 waves					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.94	1.00
Y2 (1.634)	1.64 (0.08)	0.08	0.01	0.95	1.00
Y3 (1.771)	1.77 (0.09)	0.09	0.01	0.95	1.00
Y4 (1.628)	1.63 (0.09)	0.09	0.01	0.95	1.00
Y5 (1.515)	1.51 (0.09)	0.09	0.01	0.94	1.00
Y6 (1.303)	1.29 (0.13)	0.12	0.02	0.95	1.00
Last 7 waves					
Y6 (1.303)	1.22 (0.09)	0.10	0.01	0.89	1.00
Y7 (1.214)	1.23 (0.07)	0.07	0.01	0.94	1.00
Y8 (1.144)	1.14 (0.06)	0.06	0.00	0.95	1.00

Y9 (1.183)	1.18 (0.07)	0.06	0.00	0.94	1.00
Y10 (0.935)	0.92 (0.05)	0.05	0.00	0.93	1.00
Y11 (0.742)	0.74 (0.05)	0.05	0.00	0.95	1.00
Y12 (0.956)	0.94 (0.08)	0.08	0.01	0.93	1.00

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 40

Simulation results for the Class 1 factor means across the five assessment configurations, $N = 500$.

Class 1 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.06
Linear (5.420)	5.40 (0.36)	0.34	0.13	0.95	1.00
Quadratic (-5.313)	-5.31 (0.72)	0.71	0.52	0.94	1.00
Cubic (1.643)	1.65 (0.41)	0.42	0.17	0.95	0.98
2-yr intervals					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.05
Linear (5.420)	5.41 (0.52)	0.50	0.27	0.94	1.00
Quadratic (-5.313)	-5.33 (1.26)	1.24	1.58	0.95	0.99
Cubic (1.643)	1.67 (0.79)	0.77	0.62	0.94	0.58
Uneven intervals					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.05
Linear (5.420)	5.41 (0.42)	0.41	0.18	0.93	1.00
Quadratic (-5.313)	-5.33 (0.93)	0.94	0.87	0.95	1.00
Cubic (1.643)	1.67 (0.55)	0.56	0.30	0.96	0.84
First 6 waves					
Intercept (0.003)	0.00 (0.01)	0.01	0.00	0.96	0.05
Linear (5.420)	5.46 (0.98)	0.96	0.96	0.93	1.00
Quadratic (-5.313)	-5.62 (4.91)	4.86	24.14	0.93	0.21
Cubic (1.643)	2.05 (5.99)	5.98	35.94	0.94	0.08
Last 7 waves					
Intercept (0.003)	-0.94 (0.62)	0.47	1.28	0.41	0.58
Linear (5.420)	0.71 (4.92)	4.87	46.35	0.71	0.09
Quadratic (-5.313)	-0.75 (19.74)	18.65	409.51	0.91	0.08
Cubic (1.643)	4.12 (21.65)	19.63	473.76	0.92	0.08

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 41

Simulation results for the Class 2 factor means across the five assessment configurations, $N = 500$.

Class 2 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (0.963)	0.96 (0.04)	0.04	0.00	0.94	1.00
Linear (4.533)	4.51 (1.05)	1.03	1.10	0.94	0.99
Quadratic (-4.836)	-4.78 (2.10)	2.13	4.38	0.95	0.63
Cubic (1.603)	1.57 (1.20)	1.24	1.44	0.95	0.22
2-yr intervals					
Intercept (0.963)	0.96 (0.04)	0.04	0.00	0.94	1.00
Linear (4.533)	4.57 (1.50)	1.52	2.26	0.94	0.86
Quadratic (-4.836)	-4.91 (3.55)	3.71	12.60	0.95	0.27
Cubic (1.603)	1.64 (2.20)	2.32	4.85	0.96	0.09
Uneven intervals					
Intercept (0.963)	0.96 (0.04)	0.04	0.00	0.94	1.00
Linear (4.533)	4.52 (1.22)	1.23	1.49	0.94	0.94
Quadratic (-4.836)	-4.84 (2.74)	2.82	7.49	0.95	0.43
Cubic (1.603)	1.62 (1.61)	1.68	2.60	0.94	0.14
First 6 waves					
Intercept (0.963)	0.96 (0.04)	0.04	0.00	0.95	1.00
Linear (4.533)	4.56 (3.04)	2.89	9.22	0.93	0.37
Quadratic (-4.836)	-5.04 (15.42)	14.63	237.42	0.93	0.08
Cubic (1.603)	1.88 (18.90)	17.99	356.54	0.93	0.07
Last 7 waves					
Intercept (0.963)	0.69 (0.41)	0.35	0.24	0.67	0.64
Linear (4.533)	0.81 (2.56)	2.44	20.38	0.53	0.06
Quadratic (-4.836)	-1.70 (9.67)	9.48	103.15	0.90	0.06
Cubic (1.603)	2.51 (10.26)	10.03	105.90	0.94	0.06

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 42

Simulation results for the Class 3 factor means across the five assessment configurations, $N = 500$.

Class 3 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (2.021)	2.02 (0.03)	0.03	0.00	0.93	1.00
Linear (1.863)	1.79 (0.99)	0.95	0.98	0.94	0.48
Quadratic (-1.491)	-1.38 (2.02)	1.98	4.07	0.96	0.12
Cubic (0.010)	-0.05 (1.17)	1.16	1.36	0.97	0.04
2-yr intervals					
Intercept (2.021)	2.02 (0.03)	0.03	0.00	0.93	1.00
Linear (1.863)	1.80 (1.41)	1.41	1.99	0.94	0.27
Quadratic (-1.491)	-1.45 (3.47)	3.46	12.00	0.94	0.08
Cubic (0.010)	0.00 (2.17)	2.17	4.69	0.94	0.06
Uneven intervals					
Intercept (2.021)	2.02 (0.03)	0.03	0.00	0.93	1.00
Linear (1.863)	1.80 (1.19)	1.13	1.41	0.95	0.40
Quadratic (-1.491)	-1.37 (2.74)	2.62	7.52	0.94	0.10
Cubic (0.010)	-0.06 (1.64)	1.57	2.70	0.95	0.05
First 6 waves					
Intercept (2.021)	2.02 (0.03)	0.03	0.00	0.93	1.00
Linear (1.863)	1.83 (2.81)	2.71	7.88	0.94	0.12
Quadratic (-1.491)	-1.65 (14.12)	13.75	199.09	0.95	0.05
Cubic (0.010)	0.33 (17.32)	16.92	299.50	0.95	0.05
Last 7 waves					
Intercept (2.021)	1.99 (0.41)	0.35	0.17	0.76	0.97
Linear (1.863)	0.66 (4.31)	2.54	19.96	0.89	0.08
Quadratic (-1.491)	-1.24 (16.69)	9.92	278.06	0.93	0.09
Cubic (0.010)	-0.14 (17.39)	10.59	301.79	0.93	0.07

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 43

Simulation results for the Class 4 factor means across the five assessment configurations, $N = 500$.

Class 4 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (3.262)	3.26 (0.03)	0.03	0.00	0.93	1.00
Linear (-2.191)	-2.20 (1.08)	1.04	1.17	0.93	0.57
Quadratic (3.049)	2.97 (2.23)	2.14	4.98	0.93	0.30
Cubic (-1.610)	-1.54 (1.30)	1.25	1.69	0.94	0.26
2-yr intervals					
Intercept (3.262)	3.26 (0.03)	0.03	0.00	0.94	1.00
Linear (-2.191)	-2.23 (1.64)	1.50	2.67	0.92	0.34
Quadratic (3.049)	3.10 (3.96)	3.68	15.63	0.93	0.17
Cubic (-1.610)	-1.63 (2.45)	2.31	5.98	0.93	0.14
Uneven intervals					
Intercept (3.262)	3.26 (0.03)	0.03	0.00	0.93	1.00
Linear (-2.191)	-2.20 (1.24)	1.23	1.53	0.95	0.45
Quadratic (3.049)	2.94 (2.90)	2.83	8.42	0.94	0.20
Cubic (-1.610)	-1.51 (1.76)	1.69	3.09	0.93	0.16
First 6 waves					
Intercept (3.262)	3.26 (0.03)	0.03	0.00	0.93	1.00
Linear (-2.191)	-2.13 (2.91)	2.88	8.48	0.94	0.13
Quadratic (3.049)	2.62 (15.22)	14.54	231.28	0.94	0.08
Cubic (-1.610)	-1.12 (19.02)	17.88	361.43	0.92	0.07
Last 7 waves					
Intercept (3.262)	3.28 (0.43)	0.34	0.18	0.75	0.99
Linear (-2.191)	0.98 (3.18)	2.93	20.14	0.71	0.09
Quadratic (3.049)	-3.42 (12.65)	11.38	201.62	0.90	0.08
Cubic (-1.610)	0.46 (13.63)	12.19	189.64	0.93	0.06

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 44

Simulation results for the Class 5 factor means across the five assessment configurations, $N = 500$.

Class 5 Factor Mean (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Intercept (4.785)	4.79 (0.05)	0.05	0.00	0.92	1.00
Linear (-9.614)	-9.61 (1.60)	1.49	2.54	0.92	1.00
Quadratic (13.969)	13.86 (3.42)	3.08	11.65	0.91	0.97
Cubic (-6.580)	-6.48 (2.01)	1.80	4.05	0.90	0.91
2-yr intervals					
Intercept (4.785)	4.79 (0.05)	0.05	0.00	0.92	1.00
Linear (-9.614)	-9.59 (2.37)	2.20	5.63	0.92	0.97
Quadratic (13.969)	13.87 (5.83)	5.42	33.95	0.92	0.68
Cubic (-6.580)	-6.50 (3.62)	3.41	13.08	0.92	0.50
Uneven intervals					
Intercept (4.785)	4.79 (0.05)	0.05	0.00	0.92	1.00
Linear (-9.614)	-9.55 (1.80)	1.76	3.23	0.94	1.00
Quadratic (13.969)	13.73 (4.14)	4.05	17.19	0.92	0.90
Cubic (-6.580)	-6.41 (2.50)	2.42	6.29	0.93	0.76
First 6 waves					
Intercept (4.785)	4.79 (0.05)	0.05	0.00	0.92	1.00
Linear (-9.614)	-9.35 (4.40)	4.16	19.36	0.93	0.63
Quadratic (13.969)	12.79 (22.24)	20.97	494.83	0.93	0.12
Cubic (-6.580)	-5.48 (27.55)	25.76	758.52	0.92	0.08
Last 7 waves					
Intercept (4.785)	4.92 (0.68)	0.47	0.48	0.79	1.00
Linear (-9.614)	0.43 (6.12)	5.23	138.25	0.38	0.11
Quadratic (13.969)	-2.07 (24.54)	20.30	858.25	0.78	0.10
Cubic (-6.580)	-3.28 (26.74)	21.50	724.54	0.89	0.10

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 45

Simulation results for the factor variance and covariance parameters across the five assessment configurations, $N = 500$.

Factor Variances and Covariance (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Linear (13.572)	13.48 (1.14)	1.09	1.30	0.94	1.00
Quadratic (8.401)	8.32 (0.83)	0.79	0.70	0.93	1.00
Linear-Quadratic (-10.189)	-10.11 (0.95)	0.90	0.90	0.93	1.00
2-yr intervals					
Linear (13.572)	13.53 (1.51)	1.47	2.28	0.95	1.00
Quadratic (8.401)	8.35 (1.32)	1.35	1.74	0.95	1.00
Linear-Quadratic (-10.189)	-10.15 (1.36)	1.35	1.85	0.94	1.00
Uneven intervals					
Linear (13.572)	13.47 (1.49)	1.45	2.22	0.93	1.00
Quadratic (8.401)	8.30 (1.30)	1.23	1.71	0.92	1.00
Linear-Quadratic (-10.189)	-10.09 (1.31)	1.26	1.72	0.93	1.00
First 6 waves					
Linear (13.572)	13.54 (3.74)	3.76	13.94	0.95	0.96
Quadratic (8.401)	8.58 (15.40)	15.40	236.86	0.95	0.07
Linear-Quadratic (-10.189)	-10.23 (7.20)	7.21	51.74	0.95	0.30
Last 7 waves					
Linear (13.572)	5.41 (4.65)	4.66	88.22	0.52	0.27
Quadratic (8.401)	11.83 (12.96)	13.85	179.40	0.92	0.19
Linear-Quadratic (-10.189)	-5.76 (7.57)	7.81	76.71	0.85	0.17

Note. For the data generation model and the data analysis model, the variances of the intercept and cubic factors were fixed at zero. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Table 46

Simulation results for the residual variances of the observed indicators (Y1-Y12) across the five assessment configurations, N = 500.

Residual Variances (Population Value)	Parameter Estimate: Average (SD)	Average Estimated SE	M.S.E.	95% Coverage	% Sign. Estimates
12 waves					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.94	1.00
Y2 (1.634)	1.64 (0.11)	0.11	0.01	0.95	1.00
Y3 (1.771)	1.78 (0.12)	0.12	0.01	0.96	1.00
Y4 (1.628)	1.63 (0.12)	0.11	0.01	0.93	1.00
Y5 (1.515)	1.51 (0.11)	0.11	0.01	0.93	1.00
Y6 (1.303)	1.30 (0.09)	0.09	0.01	0.94	1.00
Y7 (1.214)	1.22 (0.09)	0.09	0.01	0.95	1.00
Y8 (1.144)	1.14 (0.08)	0.08	0.01	0.94	1.00
Y9 (1.183)	1.19 (0.09)	0.08	0.01	0.93	1.00
Y10 (0.935)	0.93 (0.07)	0.07	0.00	0.95	1.00
Y11 (0.742)	0.74 (0.06)	0.06	0.00	0.94	1.00
Y12 (0.956)	0.96 (0.09)	0.09	0.01	0.94	1.00
2-yr intervals					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.94	1.00
Y3 (1.771)	1.77 (0.13)	0.12	0.02	0.94	1.00
Y5 (1.515)	1.51 (0.12)	0.12	0.01	0.94	1.00
Y7 (1.214)	1.22 (0.10)	0.10	0.01	0.96	1.00
Y9 (1.183)	1.18 (0.10)	0.10	0.01	0.94	1.00
Y11 (0.742)	0.74 (0.13)	0.14	0.02	0.96	1.00
Uneven intervals					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.95	1.00
Y2 (1.634)	1.64 (0.11)	0.11	0.01	0.94	1.00
Y3 (1.771)	1.78 (0.12)	0.13	0.01	0.96	1.00
Y4 (1.628)	1.62 (0.13)	0.13	0.02	0.92	1.00
Y8 (1.144)	1.12 (0.12)	0.12	0.02	0.92	1.00
Y12 (0.956)	0.97 (0.31)	0.30	0.10	0.94	0.89
First 6 waves					
Y1 (0.043)	0.04 (0.00)	0.00	0.00	0.94	1.00
Y2 (1.634)	1.63 (0.11)	0.11	0.01	0.94	1.00
Y3 (1.771)	1.78 (0.13)	0.13	0.02	0.95	1.00
Y4 (1.628)	1.62 (0.12)	0.12	0.01	0.95	1.00
Y5 (1.515)	1.51 (0.13)	0.12	0.02	0.92	1.00
Y6 (1.303)	1.28 (0.18)	0.17	0.03	0.93	1.00
Last 7 waves					
Y6 (1.303)	1.20 (0.14)	0.15	0.03	0.86	1.00
Y7 (1.214)	1.23 (0.10)	0.10	0.01	0.95	1.00
Y8 (1.144)	1.14 (0.09)	0.09	0.01	0.96	1.00

Y9 (1.183)	1.18 (0.09)	0.09	0.01	0.94	1.00
Y10 (0.935)	0.92 (0.07)	0.07	0.01	0.93	1.00
Y11 (0.742)	0.73 (0.07)	0.06	0.00	0.94	1.00
Y12 (0.956)	0.93 (0.12)	0.12	0.02	0.93	1.00

Note. Average Estimated S.E. = average estimated standard error. M.S.E. = mean square error. % Sign. Estimates = proportion of replications for which the parameter estimate is significantly different from zero ($p < .05$).

Figure 1. Unconditional latent growth curve model.

Figure 1. An unconditional model does not include covariates, while a conditional model does include covariates.

Figure 2. Unconditional growth mixture model.

Figure 2. An unconditional model does not include covariates, while a conditional model does include covariates.

Figure 3. Profile plots for a random sample of participants (n = 15).

Figure 4. Average trajectory from the latent growth model using 12 waves (N = 2686).

Figure 4. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month).

Figure 5. Average trajectory for each latent class from the 5-class growth mixture model using 12 waves (N = 2686).

Figure 5. Estimated means are plotted on the y-axis. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month, 3 = 19.09 drinks per month, 4 = 53.60 drinks per month, 5 = 147.41 drinks per month).

Figure 6. Estimated trajectories for each participant in Class 5 ($n = 96$) and the average trajectory across all participants in Class 5 using 12 waves.

Figure 6. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month, 3 = 19.09 drinks per month, 4 = 53.60 drinks per month, 5 = 147.41 drinks per month).

Figure 7. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus 2-year intervals (N = 2685).

Figure 7. Estimated means are plotted on the y-axis. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month, 3 = 19.09 drinks per month, 4 = 53.60 drinks per month, 5 = 147.41 drinks per month).

Figure 8. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus uneven intervals (N = 2685).

Figure 8. Estimated means are plotted on the y-axis. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month, 3 = 19.09 drinks per month, 4 = 53.60 drinks per month, 5 = 147.41 drinks per month).

Figure 9. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus the 1st six waves (N = 2685).

Figure 9. Estimated means are plotted on the y-axis. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month, 3 = 19.09 drinks per month, 4 = 53.60 drinks per month, 5 = 147.41 drinks per month).

Figure 10. Comparing average trajectories for each latent class from the 5-class GMMs using 12 waves (N = 2686) versus the last 7 waves (N = 2505).

Figure 10. Estimated means are plotted on the y-axis. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month, 3 = 19.09 drinks per month, 4 = 53.60 drinks per month, 5 = 147.41 drinks per month).

Figure 11. Percentage of participants assigned to Classes 1 to 5 in the last 7 wave configuration (legend) based on class membership using 12 waves (x-axis).

Figure 11. Percentages sum to 100% for each column. Sample sizes for Classes 1 to 5 using the last 12 waves were 1796, 211, 221, 187, and 90, respectively. One hundred eighty-one participants were not included in this analysis due to missing data.

Figure 12. Average trajectories from the latent growth models for all five assessment configurations.

Figure 12. Estimated means are plotted on the y-axis. Values on the y-axis using the logged scale can be converted to drinks per month (e.g., 0 = 0.00 drinks per month, 1 = 1.72 drinks per month, 2 = 6.39 drinks per month).

Figure 13. Sample size used in 31 applied studies.

Figure 13. The 31 studies used GMM or LCGA to analyze alcohol data collected from adolescents or young adults.

Figure 14. Profile plots of the simulated data from Class 1, Replication 1 ($n = 339$).

Simulated data from Class 1, Replication 1 ($n=339$)

Figure 14. Simulated data were generated under the experimental condition using 12 waves and $N = 500$.

Figure 15. Profile plots of the simulated data from Class 5, Replication 1 ($n = 61$).

Simulated data from Class 5, Replication 1 ($n=61$)

Figure 15 Simulated data were generated under the experimental condition using 12 waves and $N = 1500$.

Figure 16. Average parameter estimates for the mean of the intercept factor for Class 1 by assessment configuration and sample size.

Figure 16. Bars for the average standard errors are depicted. The population value was .003.

Figure 17. Average parameter estimates for the mean of the linear factor for Class 1 by assessment configuration and sample size.

Figure 17. Bars for the average standard errors are depicted. The population value was 5.42.

Figure 18. Average parameter estimates for the mean of the quadratic factor for Class 1 by assessment configuration and sample size.

Figure 18. Bars for the average standard errors are depicted. The population value was -5.313.

Figure 19. Average parameter estimates for the mean of the cubic factor for Class 1 by assessment configuration and sample size.

Figure 19. Bars for the average standard errors are depicted. The population value was 1.643.

Figure 20. Average parameter estimates for the variance of the linear factor by assessment configuration and sample size.

Figure 20. Bars for the average standard errors are depicted. The population value was 13.572.

Figure 21. Average parameter estimates for the variance of the quadratic factor by assessment configuration and sample size.

Figure 21. Bars for the average standard errors are depicted. The population value was 8.401.

APPENDIX A

Sample Mplus input file for the data generation routine and data analysis model for the experimental condition with 12 waves (Y1-Y12) and N = 1500.

TITLE: Annie's Dissertation

Monte Carlo simulation of 5-class GMM from NLSY results

DV is considered continuous - default for Generate option

NLSY POPULATION PARAMETERS WHEN:

VARIANCE INTERCEPT FIXED AT 0

VARIANCE CUBIC FIXED AT 0

DATA GENERATION

Assessment Configuration: 12 waves

Sample Size: 1500

DATA ANALYSIS

Assessment Configuration: 12 waves

MONTECARLO:

NAMES ARE y1-y12;

NOBSERVATIONS = 1500;

! Sample size;

NREPS = 500;

! Number of replications;

SEED = 12345678;

GENCLASSES = C(5);

! # classes used in data generation model;

CLASSES = C(5);

! # classes used in data analysis model;

REPSAVE = ALL;

! Saves the data from all replications;

SAVE = simdata_12w_n1500_rep*.dat;

! File storing generated data;

RESULTS = simres_12w_n1500_out.dat;

! File storing analysis results;

MODEL POPULATION:

! Data generation model;

! Values specified with an asterisk (*) are the population values;

%OVERALL%

intercpt linear quad cubic | y1@0 y2@.15 y3@.25 y4@.35 y5@.45 y6@.55
y7@.65 y8@.75 y9@.85 y10@.95 y11@1.05 y12@1.15;

[y1-y12@0];

! Intercepts of observed variables;

intercpt*0 linear*13.572

! Variances of latent factors;

quad*8.401 cubic*0;

! Variance of intercept & cubic factors *0;

linear with intercpt*0;

! Covariances among latent factors;

quad with intercpt*0;

! Covariances with intercept fixed *0;

quad with linear*-10.189;

! Covariances with cubic fixed *0;

cubic with intercpt*0;

cubic with linear*0;

cubic with quad*0;

! Residual variances;

y1*0.043 y2*1.634 y3*1.771 y4*1.628 y5*1.515 y6*1.303
y7*1.214 y8*1.144 y9*1.183 y10*0.935 y11*0.742 y12*0.956;

! Means of categorical latent variable;

[c#1*2.98];

[c#2*0.82];

[c#3*0.91];

[c#4*0.76];

! Means of latent factors by class;

%c#1%

[intercpt*0.003 linear*5.42 quad*-5.313 cubic*1.643];

%c#2%

[intercpt*0.963 linear*4.533 quad*-4.836 cubic*1.603];

%c#3%

[intercpt*2.021 linear*1.863 quad*-1.491 cubic*0.01];

%c#4%

[intercpt*3.262 linear*-2.191 quad*3.049 cubic*-1.61];

%c#5%

[intercpt*4.785 linear*-9.614 quad*13.969 cubic*-6.58];

ANALYSIS:

TYPE = MIXTURE;

MITERATIONS = 5000; ! # iterations for the EM algorithm;

MODEL:

! Describes the analysis model;

! Values specified with an asterisk (*) are used as starting values;

%OVERALL%

intercpt linear quad cubic | y1@0 y2@.15 y3@.25 y4@.35 y5@.45 y6@.55
y7@.65 y8@.75 y9@.85 y10@.95 y11@1.05 y12@1.15;

intercpt@0 linear*13.572

! Variances of latent factors;

quad*8.401 cubic@0;

! Variance of intercept & cubic factors @0;

linear with intercpt@0;

! Covariances among latent factors;

quad with intercpt@0;

! Covariances with intercept fixed @0;

quad with linear*-10.189;

! Covariances with cubic fixed @0;

cubic with intercpt@0;


```
cubic with linear@0;  
cubic with quad@0;
```

```
! Residual variances;  
y1*0.043 y2*1.634 y3*1.771 y4*1.628 y5*1.515 y6*1.303  
y7*1.214 y8*1.144 y9*1.183 y10*0.935 y11*0.742 y12*0.956;
```

```
! Means of latent factors by class;  
%c#1%  
[intercpt*0.003 linear*5.42 quad*-5.313 cubic*1.643];
```

```
%c#2%  
[intercpt*0.963 linear*4.533 quad*-4.836 cubic*1.603];
```

```
%c#3%  
[intercpt*2.021 linear*1.863 quad*-1.491 cubic*0.01];
```

```
%c#4%  
[intercpt*3.262 linear*-2.191 quad*3.049 cubic*-1.61];
```

```
%c#5%  
[intercpt*4.785 linear*-9.614 quad*13.969 cubic*-6.58];
```

```
OUTPUT: tech9;
```

APPENDIX B

Sample Mplus input file for the data analysis model for the experimental condition with two-year intervals (Y1-Y12) and N = 1500.

TITLE: Annie's Dissertation

Monte Carlo simulation of 5-class GMM from NLSY results

DV is considered continuous - default for Generate option

NLSY POPULATION PARAMETERS WHEN:

VARIANCE INTERCEPT FIXED AT 0

VARIANCE CUBIC FIXED AT 0

DATA GENERATION - DONE PREVIOUSLY, SAVED SIMULATED DATA

Assessment Configuration: 12 waves

Sample Size: 1500

DATA ANALYSIS

Assessment Configuration: Two-year intervals

DATA: TYPE = MONTECARLO;

FILE is simdata_12w_n1500_replist.dat;

VARIABLE:

names are y1-y12 c;

usevariables are y1 y3 y5 y7 y9 y11;

CLASSES = C(5);

ANALYSIS:

TYPE = MIXTURE;

MITERATIONS = 5000; ! # iterations for the EM algorithm;

MODEL:

! Describes the analysis model;

! Values specified with an asterisk (*) are used as starting values;

%OVERALL%

intercept linear quad cubic | y1@0 y3@.25 y5@.45 y7@.65 y9@.85 y11@1.05;

intercept@0 linear*13.572

! Variances of latent factors;

quad*8.401 cubic@0;

! Variance of intercept & cubic factors @0;

linear with intercept@0;

! Covariances among latent factors;

quad with intercept@0;

! Covariances with intercept fixed @0;

quad with linear*-10.189;

! Covariances with cubic fixed @0;

cubic with intercept@0;

cubic with linear@0;

cubic with quad@0;

```
! Residual variances;  
y1*0.043 y3*1.771 y5*1.515 y7*1.214 y9*1.183 y11*0.742;
```

```
! Means of latent factors by class;  
%c#1%  
[intercpt*0.003 linear*5.42 quad*-5.313 cubic*1.643];
```

```
%c#2%  
[intercpt*0.963 linear*4.533 quad*-4.836 cubic*1.603];
```

```
%c#3%  
[intercpt*2.021 linear*1.863 quad*-1.491 cubic*0.01];
```

```
%c#4%  
[intercpt*3.262 linear*-2.191 quad*3.049 cubic*-1.61];
```

```
%c#5%  
[intercpt*4.785 linear*-9.614 quad*13.969 cubic*-6.58];
```

OUTPUT: tech9;

SAVEDATA:

```
RESULTS = simres_2yr_n1500_out.dat;      ! File storing analysis results;
```

BIBLIOGRAPHY

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders IV-Text revision*. Washington, DC: American Psychiatric Association.
- Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika*, 64, 475-494.
- Arnett, J. J. (2005). The developmental context of substance use in emerging adulthood. *Journal of Drug Issues*, 35, 235-254.
- Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42(4), 757-786.
- Bauer, D. J., & Curran, P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8, 338-363.
- Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods*, 8, 384-393.
- Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, 9, 3-29.
- Bergman, L. R., & Trost, K. (2006). The person-oriented versus the variable-oriented approach: Are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Quarterly*, 52(3), 601-632.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley & Sons, Inc.
- Borsari, B., & Carey, K. B. (2001). Peer influences on college drinking: A review of the research. *Journal of Substance Abuse*, 13, 391-424.
- Brown, E. C. (2003). *Estimates of statistical power and accuracy for latent trajectory class enumeration in the growth mixture model*. (Unpublished doctoral dissertation). University of South Florida, Tampa.
- Brown, S. A., McGue, M., Maggs, J., Schulenberg, J., Hingson, R., Swartzwelder, S., ... Murphy, S. (2008). A developmental perspective on alcohol and youths 16 to 20 years of age. *Pediatrics*, 121 Suppl. 4, S290-S310.

- Cattell, R. B. (1952). The three basic factor-analytic research designs-their interrelations and derivatives. *Psychological Bulletin*, 49(5), 499-520.
- Chassin, L., Flora, D. B., & King, K. M. (2004). Trajectories of alcohol and drug use and dependence from adolescence to adulthood: The effects of familial alcoholism and personality. *Journal of Abnormal Psychology*, 113, 483-498.
- Chassin, L., Pitts, S. C., & Prost, J. (2002). Binge drinking trajectories from adolescence to emerging adulthood in a high-risk sample: Predictors and substance abuse outcomes. *Journal of Consulting and Clinical Psychology*, 70, 67-78.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational Psychological Measurement*, 20, 37-46.
- Cohen, P. (1991). A source of bias in longitudinal investigations of change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 18-25). Washington, DC: American Psychological Association.
- Colder, C. R., Campbell, R. T., Ruel, E., Richardson, J. L., & Flay, B. R. (2002). A finite mixture model of growth trajectories of adolescent alcohol use: Predictors and consequences. *Journal of Consulting and Clinical Psychology*, 70, 976-985.
- Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: temporal design considerations. *Drug and Alcohol Dependence*, 68, S85-S96.
- Eggleston, E. P., Laub, J. H., & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology*, 20, 1-26.
- Fairlie, A. M., Wood, M. D., & Laird, R. D. (2012). Prospective protective effect of parents on peer influences and college alcohol involvement. *Psychology of Addictive Behaviors*, 26(1), 30-41.
- Fan, X., Felsovalyi, A., Sivo, S. A., & Keenan, S. C. (2002). *SAS for Monte Carlo Studies: A guide for quantitative researchers*. Cary, NC: SAS Institute Inc.
- Flaherty, B. P. (2010). The role of latent class and mixture models in substance use theory. In L. M. Scheier (Ed.), *Handbook of drug use etiology: Theory, methods, and empirical findings* (pp. 513-523). Washington, DC: American Psychological Association.

- Gentle, J. E. (2003). *Random number generation and Monte Carlo methods* (2nd ed). New York, NY: Springer.
- Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 243-259). Washington, DC: American Psychological Association.
- Goudriaan, A. E., Grekin, E. R., & Sher, K. J. (2007). Decision making and binge drinking: A longitudinal study. *Alcoholism: Clinical and Experimental Research*, 31(6), 928-938.
- Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., Dufour, M. C., & Pickering, R. P. (2006). The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. *Alcohol Research & Health*, 29(2), 79-91.
- Greenbaum, P. E., & Dedrick, R. F. (2007). Changes in use of alcohol, marijuana, and services by adolescents with serious emotional disturbance: A parallel-process growth mixture model. *Journal of Emotional and Behavioral Disorders*, 15(1), 21-32.
- Greenbaum, P. E., Del Boca, F. K., Darkes, J., Wang, C.-P., & Goldman, M. S. (2005). Variation in the drinking trajectories of freshmen college students. *Journal of Consulting and Clinical Psychology*, 73, 229-238.
- Grimm, K. J., Ram, N., & Estabrook, R. (2010). Nonlinear structured growth mixture models in Mplus and OpenMx. *Multivariate Behavioral Research*, 45(6), 887-909.
- Haviland, A. M., & Nagin, D. S. (2005). Causal inferences with group based trajectory models. *Psychometrika*, 70(3), 557-578.
- Hill, K. G., White, H. R., Chung, I.-J., Hawkins, J. D., & Catalano, R. F. (2000). Early adult outcomes of adolescent binge drinking: Person- and variable-centered analyses of binge drinking trajectories. *Alcoholism: Clinical and Experimental Research*, 24, 892-901.
- Hingson, R. W., Zha, W., & Weitzman, E. R. (2009). Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. *Journal of Studies on Alcohol and Drugs, Suppl.* 16, 12-20.
- Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11, 36-53.

- Hix-Small, H., Duncan, T. E., Duncan, S. C., & Okut, H. (2004). A multivariate associative finite growth mixture modeling approach examining adolescent alcohol and marijuana use. *Journal of Psychopathology and Behavioral Assessment*, 26(4), 255-270.
- Hoeppner, B. B., Goodwin, M. S., Velicer, W. F., Mooney, M. E., & Hatsukami, D. K. (2008). Detecting longitudinal patterns of daily smoking following drastic cigarette reduction. *Addictive Behaviors*, 33, 623-639.
- Huang, D., Brecht, M.-L., Hara, M., & Hser, Y.-I. (2010). Influences of a covariate on growth mixture modeling. *Journal of Drug Issues*, 40(1), 173-194.
- Jackson, K. M., & Sher, K. J. (2005). Similarities and differences of longitudinal phenotypes across alternate indices of alcohol involvement: A methodologic comparison of trajectory approaches. *Psychology of Addictive Behaviors*, 19, 339-351.
- Jackson, K. M., & Sher, K. J. (2006). Comparison of longitudinal phenotypes based on number and timing of assessments: A systematic comparison of trajectory approaches II. *Psychology of Addictive Behaviors*, 20, 373-384.
- Jackson, K. M., & Sher, K. J. (2008). Comparison of longitudinal phenotypes based on alternate heavy drinking cut scores: A systematic comparison of trajectory approaches III. *Psychology of Addictive Behaviors*, 22, 198-209.
- Jacob, T., Koenig, L. B., Howell, D. N., Wood, P. K., & Haber, J. R. (2009). Drinking trajectories from adolescence to the fifties among alcohol-dependent men. *Journal of Studies on Alcohol and Drugs*, 70, 859-869.
- Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2011a). *Monitoring the Future national survey results on drug use, 1975-2010. Volume I: Secondary school students*. Ann Arbor: Institute for Social Research, The University of Michigan.
- Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2011b). *Monitoring the Future national survey results on drug use, 1975-2010. Volume II: College students and adults ages 19-50*. Ann Arbor: Institute for Social Research, The University of Michigan.
- Jones, B. L., & Nagin, D. S. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods & Research*, 35(4), 542-571.
- Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2, 302-317.

- Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4), 963-974.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lanza, S. T., & Collins, L. M. (2006). A mixture model of discontinuous development in heavy drinking from ages 18 to 30: The role of college enrollment. *Journal of Studies on Alcohol*, 67, 552-561.
- Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, 52(3), 377-389.
- Lerner, R. M., Schwartz, S. J., & Phelps, E. (2009). Problematics of time and timing in the longitudinal study of human development: Theoretical and methodological issues. *Human Development*, 52, 44-68.
- Li, F., Barrera, M. J. R., Hops, H., & Fisher, K. J. (2002). The longitudinal influence of peers on the development of alcohol use in late adolescence: A growth mixture analysis. *Journal of Behavioral Medicine*, 25, 293-315.
- Li, F., Duncan, T. E., Duncan, S. C., & Acock, A. (2001). Latent growth modeling of longitudinal data: A finite growth mixture modeling approach. *Structural Equation Modeling: A Multidisciplinary Journal*, 8, 493-530.
- Li, F., Duncan, T. E., & Hops, H. (2001). Examining developmental trajectories in adolescent alcohol use using piecewise growth mixture modeling analysis. *Journal of Studies on Alcohol*, 62, 199-210.
- Li, L., & Hser, Y.-I. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research*, 46(2), 266-302.
- Littlefield, A. K., & Sher, K. J. (2010). Alcohol use disorders in young adulthood. In J. E. Grant, & M. N. Potenza (Eds.), *Young adult mental health* (pp. 292-310). New York, NY: Oxford University Press.
- Littlefield, A. K., Sher, K. J., & Wood, P. K. (2009). Is "maturing out" of problematic alcohol involvement related to personality change? *Journal of Abnormal Psychology*, 118(2), 360-374.
- Liu, C. Q. (2008). *Identification of latent groups in growth mixture modeling: A Monte Carlo study*. (Unpublished doctoral dissertation). University of Virginia, Charlottesville.
- Lo, Y., Mendell, N., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.

- Miller, T. R., Levy, D. T., Spicer, R. S., & Taylor, D. M. (2006). Societal costs of underage drinking. *Journal of Studies on Alcohol*, 67, 519-528.
- Molenaar, P.C.M., & Campbell, C.G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychology*, 18(2), 112-117.
- Mooney, C. Z. (1997). *Monte Carlo simulation*. Iowa City, IA: Sage University Press.
- Muraven, M., Collins, R. L., Shiffman, S., & Paty, J. A. (2005). Daily fluctuations in self-control demands and alcohol intake. *Psychology of Addictive Behaviors*, 19, 140-147.
- Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, 8, 369-377.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.
- Muthén, B., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal Data Analysis* (pp. 143-165). Boca Raton: Chapman & Hall/CRC Press.
- Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, 24, 882-891.
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55(2), 463-469.
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, 4, 139-157.
- Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.
- Nagin, D. S., & Tremblay, R. E. (2005). Developmental trajectory groups: Fact or a useful statistical fiction? *Criminology*, 43(4), 873-904.
- Nesselroade, J. R., & Ghisletta, P. (2003). Structuring and measuring change over the life span. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding*

human development: Dialogues with lifespan psychology (pp. 317-337).
Boston: Kluwer Academic Publishers.

- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 535-569.
- Patock-Peckham, J. A., & Morgan-Lopez, A. A. (2010). Direct and meditational links between parental bonds and neglect, antisocial personality, reasons for drinking, alcohol use, and alcohol problems. *Journal of Studies on Alcohol and Drugs*, 71, 95-104.
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(2), 287-312.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185, 71-110.
- Perkins, H. W. (2002). Surveying the damage: A review of research on consequences of alcohol misuse in college populations. *Journal of Studies on Alcohol*, Suppl. 14, 91-100.
- Ram, N., & Grimm, K. J. (2009). Growth mixture modeling: A method for identifying difference in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, 33, 565-576.
- Reinecke, J., & Seddig, D. (2011). Growth mixture models in longitudinal research. *Advances in Statistical Analysis*, 95, 415-434.
- Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods*, 8, 364-368.
- Robert, C., & Casella, G. (2010). *Introducing Monte Carlo Methods with R*. New York, NY: Springer.
- SAS Institute Inc. (2008). *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute Inc.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177.
- Sher, K. J., Grekin, E. R., & Williams, N. A. (2005). The development of alcohol use disorders. *Annual Review of Clinical Psychology*, 1(1), 493-523.
- Sher, K. J., Jackson, K. M., & Steinley, D. (2011). Alcohol use trajectories and the

ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology*, 120(2), 322-335.

Shiffman, S., Kirchner, T. R., Ferguson, S. G., & Scharf, D. M. (2009). Patterns of intermittent smoking: An analysis using ecological momentary assessment. *Addictive Behaviors*, 34(6-7), 514-519.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, 35, 137-167.

Stanley, L. R., Henry, K. L., & Swaim, R. C. (2011). Physical, social, and perceived availabilities of alcohol and last month alcohol use in rural and small urban communities. *Journal of Youth and Adolescence*, 40(9), 1203-1214.

Tolvanen, A. (2007). Latent growth mixture modeling: A simulation study. Doctoral dissertation, Department of Mathematics, University of Jyväskylä, Finland.

Tooze, J. A., Grunwald, G. K., & Jones, R. H. (2002). Analysis of repeated measures data with clumping at zero. *Statistical Methods in Medical Research*, 11, 341-355.

Tucker, J. S., Orlando, M., & Ellickson, P. L. (2003). Patterns and correlates of binge drinking trajectories from early adolescence to young adulthood. *Health Psychology*, 22, 79-87.

Wechsler, H., Lee, J. E., Kuo, M., & Lee, H. (2000). College binge drinking in the 1990s: A continuing problem – Results of the Harvard School of Public Health 1999 College Alcohol Study. *Journal of American College Health*, 48, 199-210.

White, H. R., Johnson, V., & Buyske, S. (2000). Parental modeling and parenting behavior effects on offspring alcohol and cigarette use: A growth curve analysis. *Journal of Substance Abuse*, 12(3), 287-310.

White, H. R., McMorris, B. J., Catalano, R. F., Fleming, C. B., Haggerty, K. P., & Abbott, R. D. (2006). Increases in alcohol and marijuana use during the transition out of high school into emerging adulthood: The effects of leaving home, going to college, and high school protective factors. *Journal of Studies on Alcohol*, 67, 810-822.

Widaman, K. F. (1991). Qualitative transitions amid quantitative development: A challenge for measuring and representing change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances*,

unanswered questions, future directions (pp. 204-217). Washington, DC: American Psychological Association.

Willett, J. B. (1989). Some results on reliability for the longitudinal measurement of change: Implications for the design of studies of individual growth.

Educational and Psychological Measurement, 49, 587-602.

Windle, M., Mun, E. Y., & Windle, R. C. (2005). Adolescent-to-young adulthood heavy drinking trajectories and their prospective predictors. *Journal of Studies on Alcohol*, 66, 313-322.

Windle, M., & Windle, R. C. (2006). Alcohol consumption and its consequences among adolescents and young adults. In M. Galanter (Ed.), *Alcohol problems in adolescents and young adults: Epidemiology, neurobiology, prevention, and treatment*. (pp. 67-83). New York, NY: Springer Science and Business Media, Inc.

Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.

Wu, J., Witkiewitz, K., McMahon, R. J., Dodge, K. A., & Conduct Problems Prevention Research Group. (2010). A parallel process growth mixture model of conduct problems and substance use with risky sexual behavior. *Drug and Alcohol Dependence*, 111, 207-214.

Zucker, R. A. (2006). Alcohol use and the alcohol use disorders: A developmental-biopsychosocial systems formulation covering the life course. In D. Cicchetti & D.J. Cohen (Eds.), *Developmental psychopathology: Risk, disorder and adaptation* (2nd ed., Volume 3, pp. 620-656). New York: Wiley.