Using Two-Sequence Latent Transition Analysis to Model Transitions in Adolescent Substance Use and Sexual Behavior Simultaneously

Z. Tang

The Pennsylvania State University

The Methodology Center
Technical Report Series
#02-53

College of Health and Human Development

The Pennsylvania State University

The Pennsylvania State University

The Graduate School

College of Health and Human Development

USING TWO-SEQUENCE LATENT TRANSITION ANALYSIS TO MODEL TRANSITIONS IN ADOLESCENT SUBSTANCE USE AND SEXUAL BEHAVIOR SIMULTANEOUSLY

A Thesis in

Human Development and Family Studies

by

Zhiqun Tang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2001

ABSTRACT

There are two main purposes of the present study. The first is to introduce a new method called two-sequence latent transition analysis (TS-LTA) that can model two stage-sequences simultaneously. The second is to model adolescent substance use and sexual behavior, and the relation between them, over time in a longitudinal data set—the Healthy For Life data set. Both substance use and sexual behavior are conceptualized as stage-sequences, and models are drawn to express how one sequence predicts another. The substance use sequence contains five stages from no use of any substances to use of alcohol and marijuana; the sexual behavior sequence contains three stages from no sex to safe sex or unsafe sex. Four hypotheses regarding relations between substance use and sexual behavior sequences are proposed and tested. They are the following: 1. An adolescent's stage of substance use predicts his/her concurrent stage of sexual behavior. 2. An adolescent's change in his/her stage of substance use predicts change in his/her stage of sexual behavior, especially when there is advancement into marijuana use. 3. An adolescent's stage of sexual behavior predicts his/her concurrent stage of substance use. 4. An adolescent's change in his/her stage of sexual behavior predicts change in his/her stage of substance use. A series of two-sequence models were tested, and the results were not inconsistent with the hypotheses, suggesting that substance use and sexual behavior sequences predict each other.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

It is impossible to finish this study without the warm support from many people.

My deepest thanks go to my adviser Dr. Linda M. Collins. She has been providing me with profound understanding, motivation, encouragement, patient guidance and constructive comments and suggestions throughout my studies.

My appreciation goes to Dr. Michael Rovine for his being my masters' committee member. Moreover, his deep understanding of different methods sheds light on my thinking of relations that lie in methodology.

I am particularly grateful to Dr. Joe Schafer who hands-on how to do multiple imputation in S-Plus 6.0 and helps to answer certain theoretical questions that relate to multiple imputation.

Thanks are also due to Brian Flaherty who programmed the two-sequence latent transition analysis and patiently hands-on how to use the two-sequence program and to Dr. David Lemmon who debugged the two-sequence program, so that the present study can be done on time.

I would like to thank David Barro, Dr. David Wagstaff, Stephanie Lanza, Dr. John W. Graham, Dr. Eva S. Lefkowitz, Dr. Eric Loken, Dr. Frank Lawrence, Dr. Chi-Ming Kam, and Mildred Maldonado-Molina for their help and support in their ways.

Finally, I would like to thank my parents Guogui Tang and QiongZhen Ma, for their rearing me, to thank my parents-in-law, Xingye Zhang and Shuxian Yang, for their

truly support, and my husband Yu Zhang and my daughter Sijia Zhang for their enduring love.

This work was supported by NIDA grant P50 DA10075.

Chapter 1

Introduction

There are two main purposes of the present study. The first is to introduce a new method called two-sequence latent transition analysis (TS-LTA; Flaherty & Collins, 1999) that can model two stage-sequences simultaneously. The second is to model adolescent substance use and sexual behavior, and the relation between them, over time in a longitudinal data set—the Healthy For Life data set (HFL; Moberg, Piper, Wu & Serlin, 1993).

1.1 Relations between substance use and sexual behavior

Substance use and early sexual behavior are two well-known leading risk factors affecting adolescent health and well-being in the United States (DiClemente, Hansen & Ponton, 1996). They have also been considered two major subset behaviors that are covered under Jessor's term "problem behaviors" (R. Jessor, 1998; R. Jessor & S. L. Jessor, 1977), which refers to the typical problems confronting today's adolescents. Moreover, substance use and sexual behavior among adolescents have been found to be associated with each other (DiClemente, Hansen & Ponton, 1996; Sternberg, 1999; Brooks-Gunn & Furstenberg, 1989). For example, Rosenbaum and Kandel (1990) found that prior use of alcohol, cigarettes, marijuana and other illicit substances greatly increases the risk of early sex for adolescents; the higher the stage of substance

involvement, the greater the probability of early sex for adolescents. A similar relation between substance use and sexual activity has also been demonstrated in other studies (Zabin, Hardy, Smith, & Hersch, 1986; Mott & Haurin, 1988; Mensch & Kandel, 1992; Mott, Fondell, Hu, Kowaleski-Jones & Menaghan, 1996), showing that adolescents who use one or more substances by a given age are more likely than those who do not to initiate sexual intercourse and/or become sexually active; and vice versa. Other studies have shown that students who used substances, especially those who have advanced in using illicit substances (e.g., marijuana), tended to have multiple partners at the same time and were less likely to use condoms (Graves & Leigh, 1995; Stewart, 1994).

Overall, it appears that there is an association between substance use and sexual behavior during adolescence, and it is often difficult to disentangle whether it is substance use that predicts sexual behavior or it is sexual behavior that predicts substance use. Some studies indicate that involvement in substance use increases risks of early and risky sexual behavior, especially, premarital teen pregnancy for females (Mensch & Kandel, 1992), after controlling other variables (e.g., SES, family intactness, biological maturity, conventionality and school context). Others show that escalating substance use follows earlier sexual activity (Mott & Haurin, 1988; Bingham & Crockett, 1996). Still others argue that the two types of behaviors co-occur because a general single underlying pattern of proneness/vulnerability to deviance or unconventionality is assumed to contribute to relations among problem behaviors in adolescents (Jessor & Jessor, 1977; Costa, Jessor, Donovan, & Fortenberry, 1995). A number of studies (Donovan & Jessor, 1985; Donovan, Jessor, and Costa, 1988; Farrell, Danish, & Howard, 1992) have

provided empirical support for a single common factor, consistent with Jessor and Jessor's (1977) concept about a syndrome of problem behavior in adolescence. These conclusions are from methods that typically treat substance use and sexual behavior as continuous variables.

1.2 Methodology that has been applied to questions involving substance use and sexual behavior

The main methods that have been used to address relations between adolescent substance use and sexual behavior are regression, structural equation modeling (SEM) and factor analysis (FA). Regression, SEM and FA are related but different ways to look at the same variance-covariance matrix (a correlation matrix if standardized), which is a distinct feature of these well-known methods. Regression and SEM have been used to identify risk factors for either substance use or sexual behavior. Generally, studies using regression and SEM suggest that substance use and early sexual behavior are risk factors for each other because they are often correlated (Regression; e.g., Allen, Leadbreater & Aber, 1994; Mensch & Kandel, 1992; Rodney, Mupier & O'Neal, 1997; Smith, Udry & Morris, 1985; Capaldi, Crosby, & Stoolmiller, 1996; Ensminger, 1990; Rosenbaum & Kandel, 1990; Whitbeck, Yoder, Hoyt & Conger, 1999; SEM; e.g., Aseltine, 1995; Dishion, Capaldi, & Yoerger, 1999). FA and confirmatory FA have been applied to demonstrating the concept of a syndrome of problem behavior, and substance use and sexual behavior have been found to have high loadings on one factor (e.g., Donovan & Jessor, 1985; Donovan, Jessor, & Costa, 1988; Farrell, Danish, & Howard, 1992). From

FA studies, it appears that the two behaviors are correlated because of an underlying common factor.

Other methods such as (multivariate) analysis of variance (ANOVA/MANOVA) (e.g., Bingham & Crockett, 1996; Tubman, Windle & Windle, 1996) have also been applied to the study of relations between substance use and sexual behavior. For example, Bingham and Crockett (1996) classified adolescents into three groups that experienced early, middle or late timing of first sexual intercourse, and they examined longitudinal patterns of behaviors including substance use of these three groups. The researchers report that timing of sexual intercourse is associated with level of drunkenness and substance use, and furthermore the earlier the timing of sexual intercourse, the higher is the frequency of drunkenness and substance use for the adolescents.

Regression, SEM, FA and ANOVA/MANOVA have been contributing to the scientific knowledge about the relations between substance use and sexual behavior during adolescence. Especially, in the SEM framework, two growth curves of substance use and sexual behavior may be modeled simultaneously to show relations between the two growth curves over time. However, these methods cannot answer questions with respect to changing categorical variables. For example, suppose people move from no use of any substance to use of alcohol or smoking, how does this change relate to the change in their virginity status (if any)? Furthermore, if people move even further from no use to use of marijuana, how does this change relate to the change in their virginity

status and condom use, and how is this relation different from the relation between onset of alcohol or smoking and change in virginity status?

There are two typical ways in the SEM framework to look at these issues. The first is treating substance use and sexual behavior as changing continuous variables, modeling two growth curves simultaneously, and allowing correlations between the intercepts and slopes of the two growth curves. This method answers somewhat different questions from those asked above. The second is treating one of the two variables as a categorical or grouping variable (e.g., no use vs. use of alcohol), and using multiple-group latent growth curves to model developmental trajectories of sexual behavior for multiple groups of people. In the second method, differences between the trajectories express the relation between sexual activity and substance use. For example, the no use group may have a lower level of intercept and smaller slope of sexual behavior over time, and use of alcohol group may have a higher level of intercept and steeper slope of sexual behavior over time. However, the group membership (e.g., no use vs. use of alcohol) is not allowed to change over time. In short, new methods are needed to answer the above questions and the like.

ANOVA/MANOVA works similarly to the second method in the SEM framework discussed above. In other words, the group membership in ANOVA/MANOVA is not considered changing over time, either. Further, although it deals with whether any group differences are present, ANOVA/MANOVA is not a latent analysis and a measurement model; it cannot adjust for measurement error or speak to the issue of the predicting direction between grouping variables and characteristics of the

groups (e.g., substance involvement) classified by the grouping variables (e.g., the onset of sexual intercourse: early, middle vs. late sexual initiators; Bingham & Crockett, 1996).

Moreover, based on a correlation coefficient, the basis of regression, SEM, and FA, if A predicts B holds, then B predicts A also holds. So it is impossible to disentangle which event predicts which with these methods. The TS-LTA provides a different look at the relations of two changing categorical variables and can be used to disentangle this question in some sense by using conditional probability (explained in **1.3** and **2.4**).

1.3 Significance of Two-Sequence Analysis

Although sometimes it is useful to consider people's behavior along a continuum (e.g., increase in use of alcohol over time during adolescence), it can also be useful to conceptualize substance use or sexual behavior as a categorical process. For example, an adolescent who has never tried any kind of substance may be qualitatively different from one who has tried cigarettes only once, or an adolescent who has never had sexual intercourse may be qualitatively different from one who has experienced sexual intercourse even one time. Further, categorizing people may be needed in real life (e.g., for diagnosis, placement, prevention and treatment purposes).

Some researchers argue that models involving sequential stages, rather than a continuum, can better depict the advancement in substance use for adolescents. These models are known as stage-sequential models (Kandel & Faust, 1975; Kandel, 1989; Kandel, Yamaguchi & Chen, 1992; Kandel & Yamaguchi, 1993). A general pattern has been found that includes the following stages: no use of any kind of substance, use of

alcohol and/or cigarettes, and then use of marijuana and/or other illicit substances (e.g., cocaine). Latent Transition Analysis (LTA; Collins & Wugalter, 1992) has been applied to testing such stage sequential models for substance use (e.g., Graham, Collins, Wugalter, Chung, & Hansen, 1991; Collins, Graham, Long & Hansen, 1994; Collins, Graham, Rousculp, & Hansen, 1997; Collins, Hyatt, & Graham, 1998) using various data sets such as the Adolescent Alcohol Prevention Trial (AAPT; Graham, Rohrbach, Hanson, Flay & Johnson, 1989), The National Longitudinal Survey of Youth (NLSY; Frankel, McWilliams & Spencer, 1983), and The National Longitudinal Study of Adolescent Health (Add Health; Resnick, et al., 1997). LTA is a measurement model, involving latent quantities (e.g., substance use status) with multiple manifest indicators (e.g., items in a survey). Estimated parameters of LTA models not only tell how strong measurements are, but also depict how people are distributed in different stages and how stable or changeable stage membership is. However, no study has been done to consider sexual behavior and substance use jointly in a stage-sequential scenario.

Two-sequence latent transition analysis (TS-LTA; Flaherty & Collins, 1999) is an exciting extension of LTA, measuring change in two stage-sequences simultaneously. Two-sequence analysis adds new features to LTA to allow researchers to test theories about direct relations between substance use and sexual behavior over time, treating the two behaviors as changing categorical variables. First, with this method, substance use and sexual behavior can be treated as two stage-sequences that unfold independently over time. In other words, these models not only categorize people in different stages, but also model change over time in the two different stage sequences (e.g., from no use of any

substances to use of alcohol, from no sex to having sex without using condoms). Second, hypotheses regarding whether or not current stages and/or change in stages in the predictor sequence (e.g., substance use) relate to current stages and/or change in stages in the dependent or outcome sequence (e.g., sexual behavior sequence) can be tested. Whether the patterns of progression of the two sequences among adolescents unfold independently over time is reflected in the parameter estimates. Third, because conditional probabilities are not symmetric (e.g., probability of A given B may not equal probability of B given A), two-sequence analysis can estimate the asymmetric effects of relations between two sequences (a strong predictive relation may hold in one direction, but not in the other). For example, Flaherty and Collins (1999) found that the levels of tobacco use (from no use to frequent use) predicted the levels of alcohol use but the reverse did not hold, although tobacco use indicators and alcohol indicators were highly correlated.

Thus, looking at relations between variables differently, two-sequence analysis is unlike regression, SEM and FA in four ways: treating both the predictor sequence and the dependent sequence as latent categorical variables, using categorical variables as indicators, allowing the latent categorical variables to change over time, and modeling changing relations between two changing categorical variables in an asymmetric fashion so that which event precedes which can be disentangled in some sense. Detailed technical information about two-sequence analysis is provided in chapter 2.

Understanding the direct relations between changing substance use and sexual behavior among adolescents as categorical variables is of great importance for prevention

and intervention practice. Endeavors in prevention and intervention research areas have been dedicated to reducing the prevalence and incidence of these behaviors among adolescents, and researchers need to understand what and how to do it (Piper, Moberg, & King, 2000; Weissberg, Barton, & Shriver, 1997). A two-sequence model with a reasonable goodness of fit for a particular data set can provide information about who, when and what should be the foci of prevention and intervention programs for adolescent problem behaviors. For example, if substance use predicted sexual behavior, it would be more reasonable for a prevention program to be designed to target reducing substance use in addition to the delay of sexual behavior. If the use of a particular substance (e.g., alcohol or marijuana) were associated with the escalation of sexual behavior, it would be wise to reduce the availability of these substances for adolescents.

1.4 The purpose of this study

By estimating a latent contingency table underlying data (e.g., the levels of a predictor sequence by the levels of a dependent sequence), the two-sequence approach can probe into details regarding an association differently from methods that are based on a variance-covariance matrix. Using the two-sequence technique, the present study examines the theory, mainly stage-sequential models, by fitting a series of two-sequence models to a large sample of longitudinal data—the HFL. The two-sequence model quantifies concurrent and longitudinal effects of a predictor sequence on a dependent sequence. The two sequences are: substance use, made up of the following stages: No Use, Lifetime Alcohol only, Lifetime and Recent Alcohol, Lifetime and Recent Alcohol

+ Lifetime Marijuana, and Lifetime and Recent Alcohol + Lifetime and Recent Marijuana), and sexual behavior, made up of the following stages: No Sex, Safe Sex, and Unsafe Sex. These sequences are depicted in *Figure 1–1*(notation is corresponding to the stages respectively), where substance use is conceptualized as the predictor sequence and sexual behavior is the dependent sequence in Model A.

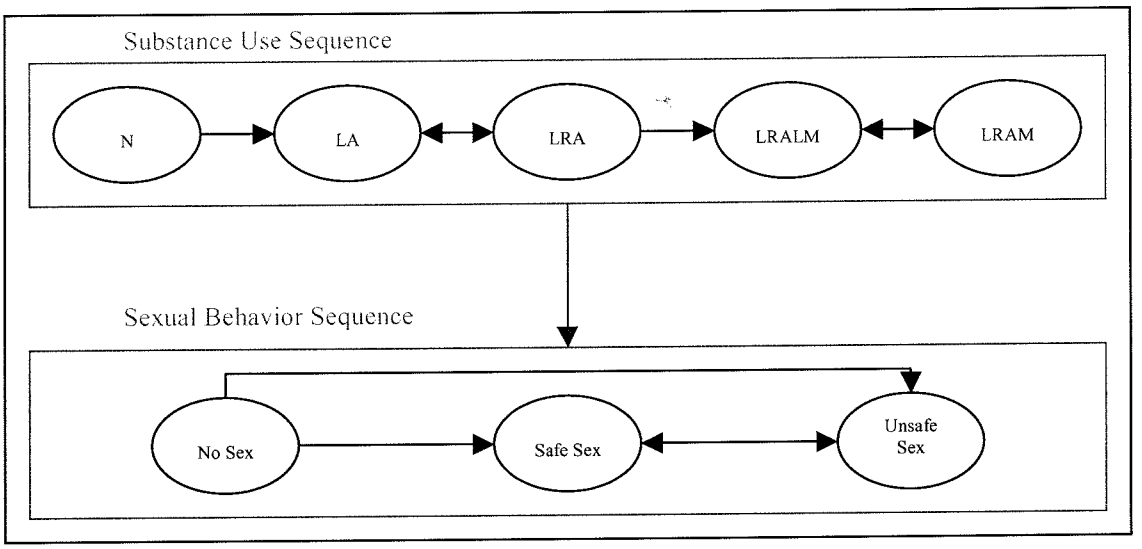

Figure 1–1: Model A—The Substance Use Sequence Predicts The Sexual Behavior Sequence

The model in *Figure 1–1* depicts the two stage sequences in a way that shows adolescents progress to more advanced levels of substance use and initiate sexual behavior over time. The model also presents a clear delineation of the influences of individuals' memberships in one stage sequence on their memberships in another stage sequence: a) the single-headed arrows within each sequence represent one-way movements between stages, b) the double-headed arrows within each sequence represent bidirectional transitions between stages, and c) the arrow that goes from one sequence to the other shows sexual behavior as a dependent or predicted or outcome sequence. The analyses can address the questions of whether or not there are underlying sequential stages with respect to substance use and sexual behavior, whether or not an adolescent's level of substance use predicts his/her concurrent level of sexual behavior, whether or not an adolescent's change in his/her level of substance use predicts change in his/her level of sexual behavior, and whether or not the reverse processes hold (See *Figure 1–2* Model B), in other words, whether the sexual behavior sequence also predicts the substance use sequence.

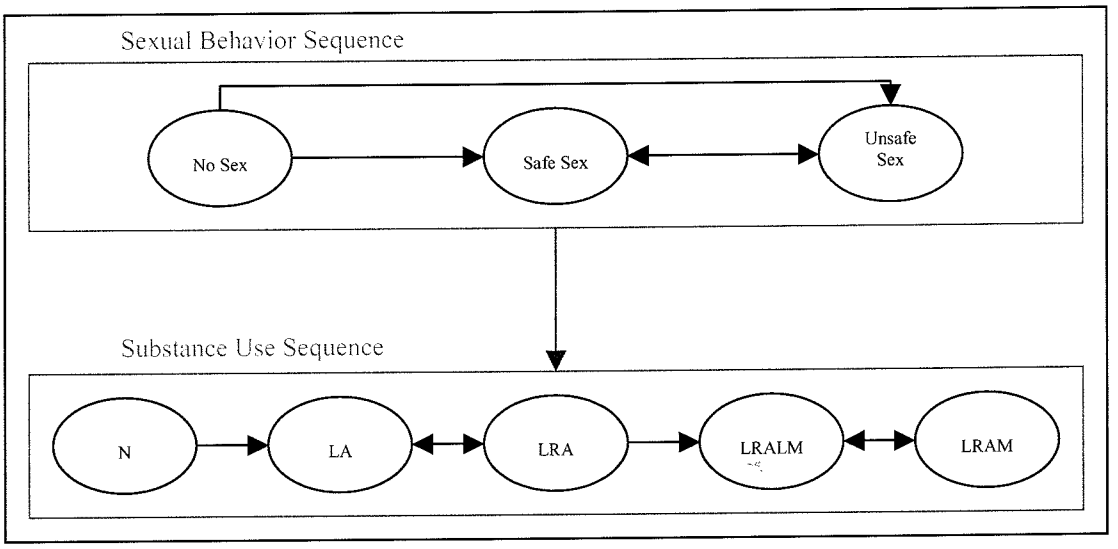

Figure 1–2: Model B—The Sexual Behavior Sequence Predicts The Substance Use Sequence

Informed by the studies reviewed above on relations between substance use and sexual behavior, four hypotheses will be tested with the HFL data:

1. An adolescent's stage of substance use predicts his/her concurrent stage of sexual behavior.
2. An adolescent's change in his/her stage of substance use predicts change in his/her stage of sexual behavior, especially when there is advancement into marijuana use.
3. An adolescent's stage of sexual behavior predicts his/her concurrent stage of substance use.
4. An adolescent's change in his/her stage of sexual behavior predicts change in his/her stage of substance use.

The hypothesis 1 and 2 relate to Model A, the hypothesis 3 and 4 relate to Model B. More detail will be given about these hypotheses in the fourth section of Chapter 2, after more information about the two-sequence latent transition model has been presented.

Chapter 2

Methods

In this chapter, the subjects and measures are described. Then, the mathematical model of TS-LTA is introduced, and the parameters in a two-sequence model are interpreted. Finally, the analytical procedures are illustrated, where the meaning of conditional probabilities and how constraints imposed on conditional probabilities relate to hypothesis testing are explained, and how missing data are handled is briefly described.

2.1 Subjects

Funded by NIDA, the Healthy for Life (HFL; Moberg, Piper, Wu & Serlin, 1993; Piper, Moberg & King, 2000) program was designed to promote health for early and middle adolescents, specifically, to influence their behavior in many areas including substance use and sexual behavior. In 1988, a sample of 2483 6^{th} grade students attending 21 Wisconsin middle schools were assessed and followed until the study ended in 1992, when most of these students were in 10^{th} grade. Thus, HFL is a longitudinal data set that contains five waves of data. The ratio of female students to male students in this data set is about fifty-to-fifty (50.1% females and 49.1 males). The majority of the sample was white (92.2%), and only a few of them (6.3%) were mixed, native-Americans, Asian Americans, Hispanic Americans, African Americans and others.

Two waves of HFL data were used in the present study—wave three and wave four. Data regarding sexuality were not collected until wave three when students were in grade eight, so wave three is the earliest wave that contains data about adolescent sexual behavior. The overall attrition rate was approximately 14% in wave three and 20% in wave four (refer to Moberg, Piper and Wu (1993) for more information about the project and subjects).

2.2 Measures

Lifetime and/or past month substance use have been used in prior studies using LTA (e.g., Graham, Collins, Wugalter, Chung, & Hansen, 1991; Collins, Graham, Long & Hansen, 1994; Collins, Hyatt, & Graham, 1998). In the present study, both kinds of measures were used as manifest indicators for Model A and Model B that were diagramed in *Figure 1-1* and *Figure 1-2* in Chapter 1. Recall that Model A is a hypothesized two-sequence model where the substance use sequence predicts the sexual behavior sequence and that Model B is a hypothesized two-sequence model where the sexual behavior sequence predicts the substance use sequence. The lifetime measures were used for measuring the latent statuses where developments over time were cumulative (represented by the single-headed arrows) and there were no backwards transitions (e.g., between no use and lifetime alcohol use). The past month measures were for latent statuses that are somewhat unstable (reflected in double-headed arrows), indicating that there were developmental reversals or fluctuations over time (e.g., between safe sex and unsafe sex, between lifetime alcohol and lifetime alcohol + recent alcohol).

The primary instrument for the HFL data set was the annual questionnaire, administered in classroom settings during regular classroom hours by research staff from local universities. Students were asked to answer the questions honestly. For the substance use sequence in Model A, four items about lifetime and past month alcohol and marijuana use were used as the manifest indicators for each measurement occasion:

1. How many times in **your whole life** have you used beer, wine or liquor?
2. How often in **the past month** did you use beer, wine or liquor?
3. How many times in **your whole life** have you used marijuana?
4. How often in **the past month** did you use marijuana?

These four items were originally measured on a 5-point scale, where 1= not at all, 2 = 1-2 times, 3 = 3-4 times, 4 = 5-8 times, and 5 = 9 or more times. After their frequency distributions had been checked, these items were recoded as two-category variables in the present study. The data that were originally coded as 1 stayed the same, meaning "no use"; the data that were originally coded as 2, 3, 4, or 5 were recoded as 2, standing for "have used"; the missing data were recoded as 0.

For the sexual sequence, three items about sexual intercourse and condom use were used as the manifest indicators for each measurement occasion:

1. Have you ever had sexual intercourse (made love, gone all the way)?
2. How many times in your life, if ever, have you had sexual intercourse?
3. If you have sexual intercourse, how often do you or your partner use a condom ("rubber")?

The first sexual intercourse item was measured on a 4-point scale, where 0= No, never, 1 = Yes, once, 2 = Yes, more than once, and 3 = not sure. The data that were originally coded as 0 were recoded as 1, meaning "no sex"; the data that were originally coded as 1, 2, or 3 were recoded as 2, standing for "have had sex"; the missing data were recoded as 0. The second sexual intercourse item was measured on a 5-point scale, where 0= Never, 1 = 1 time, 2 = 2-4 times, 3 = 5-9 times, and 4=10 or more times. This item was recoded in a similar way as the first one so that there were also two response categories, namely, "no" or "yes", and missing data were recoded as 0. For the item measuring condom use, the original 4-point scale was changed to three response categories based on an understanding of safe sex and unsafe sex. Then, 1 stands for no sex, 2 represents safe sex (always use a condom), 3 is for unsafe sex (combined "never use a condom" and "sometimes use a condom"), and 0 is missing data.

A benefit of recoding the variables based on their distributions is that it helps to reduce measurement error due to possible scaling problems. For example, measurement errors that relate to a 5-point scale, where 0= Never, 1 = 1 time, 2 = 2-4 times, 3 = 5-9 times, and 4=10 or more times, may be higher than "no" versus "have any".

2.3 The two-sequence mathematical model

In a two-sequence model, one sequence that contains different stages or levels is treated as a predictor, and another one that involves different stages or levels of other kinds is treated as the predicted or dependent or outcome sequence. The stages or levels are termed *latent statuses*, consistent with the terminology in LTA (Collins & Wugalter,

1992). Both sequences are considered as a latent variable, measured by multiple categorical manifest indicators (e.g., items in a survey). Latent status membership in the dependent sequence is conditioned on latent status membership in the predictor sequence.

For the purpose of the illustration, the model diagramed in *Figure 1-1* in Chapter 1 is used as an example. There are two-time measurement occasions in the present study. Let X denote latent status membership in the predictor sequence (also called the independent sequence) and Y denote latent status membership in the dependent sequence at time t (also called the outcome sequence). X' and Y' denote latent status membership in the same sequences at time t+1 (the notation is consistent throughout the paper). In Model A, the stages of substance use described in *Figure 1-1* are treated as the predictor sequence and levels of sexual behavior are the dependent sequence (e.g., never have sexual intercourse (Y_0=no sex), have had sexual intercourse and always use condoms (Y_1=safe sex), and have had sexual intercourse but not always or never use condoms (Y_2=unsafe sex)). Similarly, X_0 denotes the never tried any substances latent status at time t, and X_0' indicates membership in the same latent status at time t+1. For Model A, X contains five stages X_0-X_4, corresponding to N, LA, LRA, LRALM, and LRAM, and Y contains three stages Y_0, Y_1, and Y_2. It is also easy to denote Model B in this way, where X contains three stages X_0-X_2 and Y contains five stages Y_0-Y_4 (all notation for this study were given in Appendix A).

Let R represent a response pattern of $\{f, g, h, i, j, k, l, f', g', h' i', j', k', l'\}$, where there are answers to 4 items measuring the substance use sequence (items for a predictor sequence must be put first) and 3 items measuring the sexual sequence at both time t and

time t+1 (those with the prime symbol " "; time 1 must be put first and then time 2 and so on), and each item may have different response categories (e.g., the item i and i' has 1, …, I response categories, similar for the others). All items must be in the exact same order at different times. Then a two-sequence model attempts to reproduce the probability of observed response patterns P(R), conceptually resembling SEM's reproducing a variance-covariance matrix.

In order to reproduce P(R), two-sequence analysis uses the expectation-maximization (EM; Dempster, Laird, & Rubin, 1977) algorithm to find the maximum likelihood estimates for all parameters. The parameters include P(X), P($Y|X$), P($X'|XY$), and P($Y'|XYX'$).

P(X)

These parameters are the unconditional probability of being in latent status X at time t. Thus P(X) parameters are equivalent to δ (delta) parameters at time one in LTA. In Model A, the substance sequence is the predictor sequence and there are five levels of substance use in this study, so there will be five of these parameters for the five latent statuses. For example, P(X_0) is the probability that a randomly selected person from the sample will be in the never tried any substances latent status. So, P(X_0) is also the proportion of people in the never tried any substances latent status at time t.

P($Y|X$)

These parameters are the conditional probability of being in latent status Y at time t conditional on membership in latent status X at time t. For example, the probability of being in the latent status Y_1—have had sexual intercourse and always use condoms (the

safe sex group) at time t conditional on being a member of tried alcohol but without recent use of alcohol (X_1) is denoted by $P(Y_1|X_1)$. Because there are five levels in the substance use sequence and three levels in the sexual behavior sequence, there will be a total of fifteen of these concurrent parameters (see *Table 2–1*). These parameters represent the concurrent relation between the dependent and predictor sequences. If the probability of latent status membership in the predicted series varies considerably depending on latent status membership in the predictor series, this suggests that X predicts Y. This is also an important point that is used in hypothesis testing regarding conditional probabilities (coming soon in section **2.4**).

Table 2–1: The $P(Y|X)$ parameters

$P(Y\|X)$	Y_0	Y_1	Y_2
X_0	$P(Y_0\|X_0)$	$P(Y_1\|X_0)$	$P(Y_2\|X_0)$
X_1	$P(Y_0\|X_1)$	$P(Y_1\|X_1)$	$P(Y_2\|X_1)$
X_2	$P(Y_0\|X_2)$	$P(Y_1\|X_2)$	$P(Y_2\|X_2)$
X_3	$P(Y_0\|X_3)$	$P(Y_1\|X_3)$	$P(Y_2\|X_3)$
X_4	$P(Y_0\|X_4)$	$P(Y_1\|X_4)$	$P(Y_2\|X_4)$

<u>$P(X'|XY)$</u>

These parameters are the probability of being in latent status X at time t+1 conditional on the pair of latent status memberships at time t. In other words, the latent status memberships at time t of both the predictor and the dependent sequences are used to predict membership in the predictor sequence at t+1. An example of these parameters is $P(X_4'|X_0Y_0)$, the probability of being in the latent status of life time and recent use of both alcohol and marijuana at time t+1 (X_4') conditional on being in the latent statuses of

no use of substances and never have had sexual intercourse at time t (X_0Y_0). For the present model, there will be 5*3*5=75 of these parameters.

P($Y'|XYX'$)

These parameters represent the probability of being in latent status Y at time t+1 (i.e., Y') conditional on the set of latent status memberships X and Y at time t and X at time t+1 (i.e., X'). These estimates quantify the effects of changes in the predictor sequence on the dependent sequence. In the present study, for example, P($Y_2'|\ X_0Y_0X'$) denotes the probability of being in the latent status of unsafe sex at time t+1 (Y_2') conditional on no use of substances and never have had sexual intercourse at time t (X_0Y_0) and no use of substances at time t+1 (X_0'). There are five of these P($Y_2'|\ X_0Y_0X'$) parameters because X' has five levels, and these five parameters can give the intuition whether or not they vary depending on the five levels of substance use sequence at time t+1. If they do, then that is the indication that the onset of substance use (from X_0 at time t to X_1, X_2, X_3, or X_4 at t+1) relates to initiation of sexual intercourse from time t to time t+1 (i.e., from Y_0 at time t to Y_2 at time t+1). One more example, P($Y_2'|X_3Y_0X_4'$), reflects the probability of being in the same latent status of unsafe sex at time t+1 (Y_2') but conditional on some different memberships—the latent statuses of lifetime and recent alcohol and lifetime marijuana use (X_3) and never having sexual behavior (Y_0) at time t and the latent status of having lifetime and recent alcohol and lifetime and recent marijuana use at time t+1 (X_4'). Notice that this notation is for one parameter because of $X4'$ instead of X'. Because of five levels of the substance use sequence, three levels of

sexual behavior sequence, and the two measurement occasions, there will be a total of 5*3*5*3=225 parameters of P($Y'|XYX'$).

Let S_x and S_y denote the number of latent statuses in the predictor sequence and the dependent sequence, respectively, then the probability of having a particular response pattern is:

$$P(R)=\sum_{X=1}^{Sx}\sum_{Y=1}^{Sy}\sum_{X'=1}^{Sx}\sum_{Y'=1}^{Sy} P(X)P(Y|X)\rho_{f|xy}\,\rho_{g|xy}\,\rho_{h|xy}\rho_{i|xy}\rho_{j|xy}\rho_{k|xy}\rho_{l|xy}$$

$$P(X'|XY)P(Y'|XYX')\,\rho_{f'|x'y'}\,\rho_{g'|x'y'}\,\rho_{h'|x'y'}\rho_{i'|x'y'}\rho_{j'|x'y'}\rho_{k'|x'y'}\rho_{l'|x'y'}$$

In the present study, S_x=5 and S_y =3 in Model A, and S_x=3 and S_y =5 in Model B. It is not difficult to expand this mathematical model to more than two measurement occasions and more than seven manifest indicators, and to include a static component, in other words, to include latent classes.

2.4 Analytical procedures

A main purpose of the present study is fitting a series of models (a total of ten models) including and stemming from Model A and Model B using the two-sequence technique with the HFL data set (all the analyses were done by the TS-LTA Fortran program). The actual data analyzed by TS-LTA were response patterns, arrays of answers to the selected manifest indicators at different measurement occasions, a similar format to what is analyzed in LTA. The difference is there are two sequences now.

2.4.1 Conditional probability and parameter restrictions

Except for P(X) parameters, parameters of a two-sequence model, namely P($Y|X$), P($X'|XY$), and P($Y'|XYX'$), are conditional probabilities. For example, if we look at the marginal of *Table 2–1*, there are three levels of in the sexual behavior sequence in Model A for P($Y|X$), i.e., Y_0,Y_1, and Y_2. Within the table, there are also three Y's for each level of X. For example, P($Y_0|X_0$), P($Y_1|X_0$), and P($Y_2|X_0$). They are the probabilities in the first row in the table. P($Y_0|X_0$), P($Y_0|X_1$), P($Y_0|X_2$), P($Y_0|X_3$), and P($Y_0|X_4$) are the probabilities in the first column in the table. The probabilities of the cells in the table are probabilities that are conditional on each level of X, or simply speaking, conditional probabilities. Further, if one assumes that the probability of being in Y_0 (no sex) has nothing to do with the levels of substance use, these five probabilities in the first column should be very similar (say, 0.205, 0.213, 0.198, 0.211, and 0.207). One could make the same prediction for the five P($Y_1|X$)s and the five P($Y_2|X$). On the other hand, if the probability being in Y_0 or Y_1 (safe sex) or Y_2 (unsafe sex) does vary considerably depending on the levels of X, i.e., they are very different from each other (say, 0.954, 0.853, 0.724, 0.214, and 0.011), one can argue that Y relates to X or X predicts Y. This idea is important for hypothesis testing regarding conditional probabilities, and the two-sequence technique allows researchers to use restrictions either to force the parameters to be equal or to be freely estimated to test which model fits data better. For the examples in the above parentheses, if the numbers in the first parenthesis were the true values of the five P($Y_0|X$) parameters, then a model that force these parameters to be equal would not fit worse than a model that allows these parameters freely estimated. In contrast, if the numbers in the second

parenthesis were the true values of P($Y_0|X$) parameters, then a model that force these parameters to be equal would probably fit significantly worse than a model that allows these parameters freely estimated. This idea is carried over through all hypotheses testing in the present study. Now let us turn to illustrating which kinds of restrictions can be imposed on a two-sequence model.

Parameter restrictions may be imposed on a two-sequence model. There are two kinds of parameter restrictions. One type of restriction is when a parameter is fixed at some particular value, i.e., without estimating it. Another type of restriction is an equality constraint. That is, a set of parameters are forced to be equal to each other and form an equivalence set, which is counted as only one parameter to be estimated.

The purpose of using parameter restrictions is to enable model identification and model specification, often for the purpose of testing theoretical hypotheses as described above. Parameter restrictions can be used to reduce the number of parameters to be estimated. To identify a two-sequence model, the degrees of freedom must be positive. That is, the number of possible response patterns minus the number of parameters to be estimated and then minus one should be a positive number. If a model has a lot of parameters to be estimated and the sample does not have enough information, the model may not be identified even with positive degrees of freedom. Usually, imposing restrictions will help to solve this problem.

Parameter restrictions can be used to specify a model. For example, in Model A, backward transitions from safe sex and unsafe sex to no sex are assumed impossible. To specify this, all parameters of P($Y_0'|XY_1X'$) and P($Y_0'|XY_2X'$) must be fixed to a value of

zero in other words, for those who already had sex in grade eight, reporting that they had not ever had sex when they were in grade nine was illogical, and thus the illogical movements were restricted by the model.

Theoretical hypothesis testing can be conducted by comparing models that are identical except that one imposes constraints and one does not. For example, to test a hypothesis that gender predicts or relates to alcohol use, a simple 2x2 frequency table can be created like *Table 2–2* based on 200 people, with half females and half males. In this case, the conditional probability P(Yes|Female)=42/100=0.42, P(No|Female)=58/100=0.58; whereas P(Yes|Male)=78/100=0.78, P(No|Male)=22/100=0.22. If gender and alcohol use were independent, we would expect that P(Yes|Female)=P(Yes|Male) and P(No|Female)=P(No|Male), and the frequency should be similar to those in the parentheses. In other words, conditioning on gender (i.e., male or female) does not provide any additional information about the proportion of people who use alcohol. However, here we have P(Yes|Female) ≠ P(Yes|Male) and P(No|Female) ≠ P(No|Male). The meaning of this inequality is that females and males have different probabilities of using alcohol. To put it in another way, gender relates to or predicts alcohol use. When testing the hypothesis that gender predicts or relates to alcohol use, we use the reverse

Table 2–2: An example of conditional probability and hypothesis testing

Frequency		Alcohol		*Total*
		Yes	No	
Gender	Female	42 (60)	58 (40)	100
	Male	78 (60)	22 (40)	100

logic: forcing the two probabilities in each column to be equal (e.g., by imposing equality constraints), calling it a constrained model, comparing it with the model that has these probabilities freely estimated, and then checking whether or not the two models fit the data equally well. For this example, we would expect that the constrained model would probably fit worse than the freely estimated one because 0.42 and 0.58 for females appear to be very different from 0.78 and 0.22 for males. The idea of comparing a model that has equality constraints with a model that does not is the basic idea underlying all the analyses in the present study.

For example, to test hypothesis 1—an adolescent's stage of substance use predicts his/her concurrent stage of sexual behavior—two models were compared. The difference between the two models was that one had equality constraints (represents a null hypothesis) and one did not (represents an alternative hypothesis). The alternative-hypothesis model, Model A (for hypotheses 1 and 2), had all fifteen $P(Y|X)$ parameters freely estimated, i.e., without equality constraints (refer back to *Table 2–1*) (actually, because the conditional probability of each row in the table sums to one, only ten independent parameters were estimated for these fifteen parameters). The restricted model had three sets of equality constraints (see *Table 2–3*). As described above, the cells that have the same number form an equivalence set, which is counted as only one parameter to be estimated. So only two parameters were estimated for these three sets of parameters. Therefore, by

Table 2–3: Equality constraints for testing hypothesis 1—Model A-1

$P(Y\|X)$	Y_0	Y_1	Y_2
X_0	21	31	41
X_1	21	31	41
X_2	21	31	41
X_3	21	31	41
X_4	21	31	41

imposing the equality constraints like this, the restricted model gained eight degrees of freedom (ten minus two). Which model fits data better? The difference of G-squared from the two analyses is distributed as chi-squared with the degrees of freedom equal to the difference between the two degrees of freedom of the two models, i.e., eight in this case. A significant G-squared difference indicates that imposing equality constraints significantly reduces the model fit, so that it would be more reasonable to freely estimate the parameters at the cost of some more degrees of freedom. In other words, the alternative hypothesis should be accepted and the null should be rejected; it was improper to force these parameters to be equal, and these parameters in each column may be very different, meaning that probability in *Y* varies significantly depending on levels of *X* and thus, suggesting that *X* predicts *Y*. In other words, the stages of substance use predict the concurrent stages of sexual behavior. Other hypotheses were tested in the same way—setting up the null hypotheses that were expected to be rejected when being test against the alternative hypotheses. And again, the only difference was that the same pattern of equality constraints was imposed on different sets of parameters (please keep this in mind).

In short, if a constrained model fits equally well as an unconstrained one, then the null hypothesis cannot be rejected, and conditioning on the predictor sequence does not provide more information about the proportions of people in the dependent sequence. In this case, the predictor sequence does not appear to predict the dependent sequence. If a constrained model fits significantly worse than an unconstrained one, the null hypothesis should be rejected, and it would be reasonable to accept the alternative hypothesis and

say that the parameters may be different. In other words, conditioning on different levels of substance use, the probability in different stages of sexual behavior is different, and they should not be constrained equal. More detail about constraints in the present study is provided next.

2.4.2 Hypotheses testing: the alternative-hypothesis models and equality constraints for the null- hypothesis models

As described above, different theoretical hypotheses can be tested against the null hypotheses by imposing different constraints on the null-hypothesis models. In this section, the constraints reflecting the null hypotheses are explained one by one with respect to the four hypotheses at the end of Chapter 1.

Before testing any hypothesis, Model A, the alternative-hypothesis model for hypothesis 1 and 2, was first tested to see if it fits the HFL data. First of all, equivalence constraints were imposed on big ρ's (see *Table 2–4*). Recall that big ρ's are the probability of a particular response to each item given latent status membership in the two sequences and time, reflecting measurement error (Please refer to Chapter 1 or Appendix A for notation of the variable *X* and *Y*). The meaning of big ρ's is exactly the same as big ρ's in LTA. That is, those big ρ's that either close to one or zero are strong ρ's, indicating good measurement. The only difference is, in LTA, latent status membership is in one sequence instead of two. Usually, there are two types of measurement errors that can be identified for each indicator except for condom use item. One is the probability for those who should answer "no" according to their latent status to

answer "yes" (e.g., the probability of an adolescent in the "No Use" latent status answering "yes" to the marijuana item). Another is the probability for those who should answer "yes" according to their latent status to answer "no" (e.g., the probability of an adolescent in the "Lifetime Alcohol" latent status to answer "no" to the lifetime alcohol item). Big ρ parameters were constrained to be equal for each type of measurement error within each indicator, so that there is a single big ρ parameter to be estimated for each type of measurement error for each indicator, for a total of two parameters per indicator.

In addition, the constraints imposed on the two items measuring lifetime sexual intercourses were the same. Thus, there were ten big ρ parameters in total to be estimated for the first six indicators, that is, two per indicator except for the item of condom use (the seventh item). The condom use item had six sets of equivalence constraints due to its having three response categories (constraints for responding not always use or no use of condoms were provided separately in *Table 2–5*, it is actually *Table 2–4* continued. Otherwise, *Table 2–4* would be very big.

Table 2–4: Equality constraints for big ρ's—response categories "No" and "Yes" in Model A

Status	LifA	ReA	LifM	ReM	LifSex1	LifSex2	Condom
	Answer No						No Sex
X_0Y_0	2	3	4	5	6	6	7
X_0Y_1	2	3	4	5	10	10	8
X_0Y_2	2	3	4	5	10	10	9
X_1Y_0	12	3	4	5	6	6	7
X_1Y_1	12	3	4	5	10	10	8
X_1Y_2	12	3	4	5	10	10	9
X_2Y_0	12	13	4	5	6	6	7
X_2Y_1	12	13	4	5	10	10	8
X_2Y_2	12	13	4	5	10	10	9
X_3Y_0	12	13	14	5	6	6	7
X_3Y_1	12	13	14	5	10	10	8
X_3Y_2	12	13	14	5	10	10	9
X_4Y_0	12	13	14	15	6	6	7
X_4Y_1	12	13	14	15	10	10	8
X_4Y_2	12	13	14	15	10	10	9
	Answer Yes						Yes Sex Use Condom
X_0Y_0	22	23	24	25	16	16	17
X_0Y_1	22	23	24	25	110	110	18
X_0Y_2	22	23	24	25	110	110	19
X_1Y_0	212	23	24	25	16	16	17
X_1Y_1	212	23	24	25	110	110	18
X_1Y_2	212	23	24	25	110	110	19
X_2Y_0	212	213	24	25	16	16	17
X_2Y_1	212	213	24	25	110	110	18
X_2Y_2	212	213	24	25	110	110	19
X_3Y_0	212	213	214	25	16	16	17
X_3Y_1	212	213	214	25	110	110	18
X_3Y_2	212	213	214	25	110	110	19
X_4Y_0	212	213	214	215	16	16	17
X_4Y_1	212	213	214	215	110	110	18
X_4Y_2	212	213	214	215	110	110	19
							Yes Sex No Condom (see *Table 2–5*)

Table 2–5: Equality constraints for big ρ's—response categories "No" to condom use in Model A (*Table 2–4* continued)

X_0Y_0	X_0Y_1	X_0Y_2	X_1Y_0	X_1Y_1	X_1Y_2	X_2Y_0	X_2Y_1	X_2Y_2	X_3Y_0	X_3Y_1	X_3Y_2	X_4Y_0	X_4Y_1	X_4Y_2
27	28	29	27	28	29	27	28	29	27	28	29	27	28	29

Further, big ρ parameters were constrained equal across the two measurement occasions. Thus, there were a total of sixteen big ρ parameters to be estimated for each model in the present study. When the big ρ's are constrained equal across time, the meaning of the latent statuses remains the same across time. This forces measurement invariance across time and aids in interpretation of the estimated parameters (much like factor invariance across measurement occasions or groups in factor analysis).

The five P(X) parameters and all 15 P($Y|X$) parameters were freely estimated. Among transition parameters, P($X'|XY$) and P($Y'|XYX'$), those reflecting illogical backward transitions, were fixed to a value of zero so as to specify the single-headed arrows in Model A. In other words, none of the backward transition parameters was estimated. *Table 2–6* gives the restrictions for the set of parameters P($X'|XY$). The twenty-four cells with zeros were not estimated and their values were zeros; those with ones were freely estimated.

Table 2–6: Parameter restrictions for P($X'|XY$) in Model A

Grade 8		X_0'	X_1'	X_2'	X_3'	X_4'
X_0	Y_0	1	1	1	1	1
	Y_1	1	1	1	1	1
	Y_2	1	1	1	1	1
X_1	Y_0	0	1	1	1	1
	Y_1	0	1	1	1	1
	Y_2	0	1	1	1	1
X_2	Y_0	0	1	1	1	1
	Y_1	0	1	1	1	1
	Y_2	0	1	1	1	1
X_3	Y_0	0	0	0	1	1
	Y_1	0	0	0	1	1
	Y_2	0	0	0	1	1
X_4	Y_0	0	0	0	1	1
	Y_1	0	0	0	1	1
	Y_2	0	0	0	1	1

Similarly, all parameters of P(Y_0'|XY_1X') and P(Y_0'|XY_2X') were fixed to a value of zero. *Table 2–7* gives examples of restrictions for some P(Y'|XYX') parameters.

Table 2–7: An example of parameter restrictions for P(Y'|XYX')—P(Y'|X_4YX')

Grade 8		Grade 9	Y_0	Y_1	Y_2
X_4	Y_0	X_0	0	0	0
		X_1	0	0	0
		X_2	0	0	0
		X_3	1	1	1
		X_4	1	1	1
	Y_1	X_0	0	0	0
		X_1	0	0	0
		X_2	0	0	0
		X_3	0	1	1
		X_4	0	1	1
	Y_2	X_0	0	0	0
		X_1	0	0	0
		X_2	0	0	0
		X_3	0	1	1
		X_4	0	1	1

If Model A, with the above patterns of restrictions imposed (see *Table 2–6* and *Table 2–7*), had a reasonably good fit, and the pattern of big ρ parameter estimates (see *Table 2–4* and *Table 2–5*) was consistent with the proposed latent statuses in the two sequences, then Model A can be used as a way to summarize the relations between substance use and sexual behavior as measured by the selected seven items. Model A is the alternative-hypothesis model (for hypothesis 1 and 2) against which the corresponding null hypotheses were tested; the corresponding null-hypothesis models were identical to Model A except for the imposed constraints. Model B is another alternative-hypothesis model (for hypothesis 3 and 4) against which the corresponding null hypotheses were tested. Model B is a reversed way to look at the relations between substance use and sexual behavior as measured by the same selected seven items, and it had the same patterns of equality constraints on big ρs and the same patterns of restrictions on P(Y|X) and P(Y'|XYX') parameters, except for the shift of the predictor sequence and the dependent sequence (i.e., X is latent status membership in the

sexual behavior sequence). The corresponding null-hypothesis models that test against Model B were identical to Model B except for the imposed constraints.

In terms of hypotheses testing, all models with equality constraints were the null-hypothesis models that were expected to be rejected, so that the alternative-hypothesis models without equality constraints (i.e., Model A and Model B) can be accepted. Further, constraints reflecting specific null hypothesis testing were of the same pattern as in *Table 2–3*, but were imposed on different sets of parameters as follows:

Hypothesis 1

An adolescent's stage of substance use predicts his/her concurrent stage of sexual behavior. To test this hypothesis, the null hypothesis regarding P(*Y*|X) parameters was established by Model A-1, where the substance use sequence was regarded as the predictor sequence *X*, and equality constraints were imposed on the fifteen P(*Y*|X) parameters (i.e., *Table 2–3*). Then the fit of Model A-1 was compared to Model A, which represented the alternative hypothesis, i.e., hypothesis 1. A significantly increased G-squared was expected for Model A-1, and thus rejecting the null and accepting the alternative hypothesis were expected.

Hypothesis 2

An adolescent's change in his/her stage of substance use predicts change in his/her stage of sexual behavior, especially when there is advancement into

marijuana use. To test this hypothesis (the alternative hypothesis specified by Model B), the null hypotheses regarding P(Y|XYX') parameters were set up by Model A-2, Model A-3, Model A-4, and Model A-5, where the substance use sequence was regarded as the predictor sequence X, and equality constraints were imposed on different sets of P(Y|XYX') parameters. There were two analyses to examine the first half of hypothesis 2 (i.e., an adolescent's change in his/her stage of substance use predicts change in his/her stage of sexual behavior). In the first analysis (call it null Model A-2), the equality constraints in *Table 2–3 were* imposed on the fifteen P(Y'| X_0Y_0X') parameters; in the second analysis (call it null Model A-3), the equality constraints in *Table 2–3 were* imposed on the fifteen P(Y'| X_0Y_1X') parameters. Both ModelA-2 and ModelA-3 were tested separately against Model A using G-Squared difference as a test statistic. A significantly reduced model fit was predicted for these two constrained models compared with Model A. The use of Model A-3 was to validate the same conclusion drawn from Model A-2. To test the second half of the hypothesis (i.e., especially when there is advancement into marijuana use), another two analyses were conducted and test against each other, using Model A as the model expressing the alternative hypothesis. In one analysis (call it null Model A-4), equality constraints were imposed on the first three rows of P(Y'| X_0Y_0X') (see *Table 2–8*); in another analysis (call it null

Table 2–8: Equality constraints for testing hypothesis 2—Model A-4

P(Y'|X_0Y_0X')	Y_0	Y_1	Y_2
X_0	21	31	41
X_1	21	31	41
X_2	21	31	41
X_3	1	1	1
X_4	1	1	1

Model A-5), equality constraints were imposed on the first, the fourth, and the fifth rows of P($Y'|X_0Y_0X'$) (see *Table 2–9*). So Model A-4 and ModelA-5 had the same degrees of freedom, but a larger increase of G-Squared was predicted for Model A-5 than Model A-4 (equality constraints were imposed on no use of substance and already use of marijuana in Model A-5). This prediction was made because bigger discrepancies were expected between the first row and the last two rows than between the first row and the second and the third if we think of the corresponding latent statuses in the X sequence.

Table 2–9: Equality constraints for testing hypothesis 2—Model A-5

P($Y'|X_0Y_0X'$)	Y_0	Y_1	Y_2
X_0	21	31	41
X_1	1	1	1
X_2	1	1	1
X_3	21	31	41
X_4	21	31	41

Hypothesis 3:

An adolescent's stage of sexual behavior predicts his/her concurrent stage of substance use. To test this hypothesis (i.e., the alternative hypothesis specified by Model B), the null hypothesis was established by Model B-1, where the sexual behavior sequence was regarded as the predictor sequence X, and equality constraints were imposed on the fifteen P($Y|X$) parameters (see *Table 2–10*). If we compare Model B-1 with Model A-1 (*Table 2–3*), we see the similar

Table 2–10: Equality constraints for testing hypothesis 3—Model B-1

P($Y|X$)	Y_0	Y_1	Y_2	Y_3	Y_4
X_0	21	31	41	51	61
X_1	21	31	41	51	61
X_2	21	31	41	51	61

pattern of equality constraints in each column. The difference is that the predictor sequence in Model A-1 (the substance use sequence had five latent statuses that were in the five rows in *Table 2–3*) is now the dependent sequence in Model B-1 (the substance use sequence had five latent statuses that were in the five columns).

Hypothesis 4

An adolescent's change in his/her stage of sexual behavior predicts change in his/her stage of substance use. For the null hypothesis, the sexual behavior sequence was regarded as the predictor sequence X, and the equality constraints were imposed on sets of $P(Y'|XYX')$ parameters. Two analyses were conducted. In one analysis (call it null Model B-2), five equivalent constraints were imposed on $P(Y'|X_0Y_0X')$ parameters (see *Table 2–11*). A significantly reduced model fit was predicted for Model B-2 when it was tested against Model B (represents hypothesis 4).

Table 2–11: Equality constraints for testing hypothesis 4—Model B-2

$P(Y'\|X_0Y_0X')$	Y_0	Y_1	Y_2	Y_3	Y_4
X_0	21	31	41	51	61
X_1	21	31	41	51	61
X_2	21	31	41	51	61

To validate the conclusion, another analysis was conducted against Model B., where four equivalent constraints were imposed on $P(Y'|X_0Y_1X')$ parameters (call it null Model B-3 see *Table 2–12*).

Table 2–12: Equality constraints for testing hypothesis 4—Model B-3

$P(Y'\|X_0Y_1X')$	Y_0	Y_1	Y_2	Y_3	Y_4
X_0	0	31	41	51	61
X_1	0	31	41	51	61
X_2	0	31	41	51	61

In sum, there were ten analyses in the present study, including two analyses for testing alternative-hypothesis model fit and eight analyses for testing the null hypotheses.

2.5 Multiple Imputation for missing data

There is not a missing data procedure available in the latest TS-LTA Fortran program, but there were missing data in the response patterns, so the multiple imputation technique (MI; Schafer, 1997) was implemented in order to base TS-LTA analyses on all the available data. MI replaces each missing value by a vector of $M>1$ imputed values. Generally, there are three steps to doing MI. The first step is generating or imputing M complete data sets; the second step is analyzing the M data sets using desired methods (here the method is TS-LTA); the third step is combining the M sets of results to get the parameter estimates.

Missing data will not cause biases in parameter estimation if it is missing completely at random (MCAR) (refer to Schafer, 1997 for a good review of missing data mechanisms). In other words, MCAR missingness does not depend on observed data and missing data in the sample. However, MCAR is rare in a real data set, and if it does not hold, usually estimated parameters will be biased if listwise deletion is used. Sometimes, missing data may be reasonably assumed missing at random (MAR) if missingness depends on observed data but does not depend on missing data in the sample. For example, if people who ever used alcohol and people who did not ever use alcohol have different probabilities of having missing data on the sexual intercourse variable, the missing information on sexual intercourse may be MAR conditional on alcohol use.

Under MAR, missing data may be imputed prior to any two-sequence latent analysis using multiple imputation with CAT or MIX (Schafer's website: http://methodology.psu.edu/. CAT deals with categorical missing data, and MIX can handle both categorical and continuous missing data at the same time. If the MAR assumption is not met, in other words, if missingness is nonignorable, then all imputation methods under the MAR assumption may provide biased estimates.

Assuming MAR, analyzing properly imputed data sets and then combining the results from separate but repeated analyses using the MI combination rules of Rubin (1987) not only increases power but also takes into consideration the uncertainty due to missing data (although the latter part was not considered in this study). In addition, information from covariates that may relate to missingness could be used in an imputation model.

For the first step in the present study, the missing data in the recoded original variables at two measurement occasions and the lifetime alcohol use item at grade six (a total of 15 variables) were imputed using S-PLUS 6.0 professional release, where there is the S+MISSINGDATA library supporting model-based missing data methods described by Little and Rubin (1987). The purpose of including this lifetime alcohol use item at grade six was using it as a covariate to aid MI, but it was not used in any two-sequence model in this study. The MI procedure was implemented under the Loglinear saturated model for two reasons in the present study: one is all variables are categorical or factor variables, and another is to keep the MI model to be consistent with the TS-LTA models.

EM converged very quickly at 172 iterations, and EM also converged at similar parameter estimates using different staring values. To examine how data augmentation (DA) behaves, the autocorrelation function (ACF) was plotted against lag for the worst linear function of the parameters (Please refer to *Figure 2–1*, the scale of X axis is 1:10). The plot showed that the ACF dropped down quickly before the first 200 cycles. After that, it went up and down slightly between the two dotted lines that indicate an approximate confidence interval about zero. The plot bottomed out near four thousand cycles, indicating that the parameters were independent around that number of iterations.

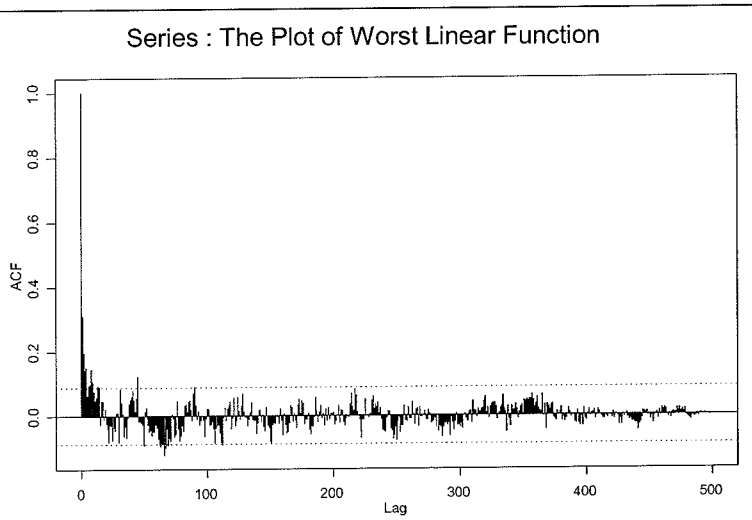

Figure 2–1: Samples of ACFs for the worst linear function of parameter estimates

To be confident that the parameters were independently drawn from the stabilized posterior distribution and being limited by the computer capacity to handle this complicated MI model, a strategy was adopted to impute data sets one by one. Under this strategy, DA was run for 5000 cycles, and the parameters of the last two runs were saved.

The DA parameter estimates in the last run were then used as the starting values to impute one data set. This process was repeated for another nine times to get a total of ten imputed data sets. Because the missing rate of the data was less than 30%, the ten imputed data sets should have more than 97% efficiency for estimated parameters.

Having obtained the ten imputed data sets, each of the ten models was run with these ten data sets. The final results for each model were the averaged parameter estimates from the ten runs.

Chapter 3

Results

In this chapter, model fit and parameter estimates were presented. The results were all consistent with the hypotheses.

3.1 Model fit for the two alternative-hypothesis models

The G-Squared for Model A (see *Figure 1-1*) was 3646.8 with 36729 degrees of freedom. The G-Squared for Model B (see *Figure 1-2*) was 3646.8 with 36729 degrees of freedom. So both alternative models had a good model fit. Further, big ρ parameters were reasonably strong. The probabilities to give a "yes" or a "no" answer to each item ranged from 0.7373 to 1.0000 for both Model A and Model B at the high end, and from 0.2627 to 0.0000 at the low end. Recall that these parameter estimates represent the probabilities of giving a "yes" or a "no" answer to the selected items conditional on latent status membership of the sampled subjects in the two sequences, and recall that big ρ's either close to one or zero is an indication of good measurement of an item (i.e., how much does answering an item match the corresponding latent status, similar to factor loadings in SEM but not identical). For example, given no use of any substance (the first three rows in the first half of *Table 3–1*), the probabilities to answer "no" to each of the substance items (the first four columns) were pretty high (0.8576-0.9993), and the probabilities to answer "yes" to each of the substance items (the first four columns) were

pretty low (0.1424-0.0007)—they can be understood as measurement errors (the first three rows in the second half of *Table 3–1*).

Table 3–1: Big ρ parameter estimates for Model A

Status	LifeA	RectA	LifM	RectM	LifSx1	LifSx2	Condom
	Answer No						No Sex
X_0Y_0	0.8576	0.9992	0.9968	0.9993	0.9920	0.9920	0.9857
X_0Y_1	0.8576	0.9992	0.9968	0.9993	0.1957	0.1957	0.1723
X_0Y_2	0.8576	0.9992	0.9968	0.9993	0.1957	0.1957	0.1919
X_1Y_0	0.0599	0.9992	0.9968	0.9993	0.9920	0.9920	0.9857
X_1Y_1	0.0599	0.9992	0.9968	0.9993	0.1957	0.1957	0.1723
X_1Y_2	0.0599	0.9992	0.9968	0.9993	0.1957	0.1957	0.1919
X_2Y_0	0.0599	0.2263	0.9968	0.9993	0.9920	0.9920	0.9857
X_2Y_1	0.0599	0.2263	0.9968	0.9993	0.1957	0.1957	0.1723
X_2Y_2	0.0599	0.2263	0.9968	0.9993	0.1957	0.1957	0.1919
X_3Y_0	0.0599	0.2263	0.2627	0.9993	0.9920	0.9920	0.9857
X_3Y_1	0.0599	0.2263	0.2627	0.9993	0.1957	0.1957	0.1723
X_3Y_2	0.0599	0.2263	0.2627	0.9993	0.1957	0.1957	0.1919
X_4Y_0	0.0599	0.2263	0.2627	0.0000	0.9920	0.9920	0.9857
X_4Y_1	0.0599	0.2263	0.2627	0.0000	0.1957	0.1957	0.1723
X_4Y_2	0.0599	0.2263	0.2627	0.0000	0.1957	0.1957	0.1919
	Answer Yes						Yes Sex Use Condom
X_0Y_0	0.1424	0.0008	0.0032	0.0007	0.0080	0.0080	0.0064
X_0Y_1	0.1424	0.0008	0.0032	0.0007	0.8043	0.8043	0.8229
X_0Y_2	0.1424	0.0008	0.0032	0.0007	0.8043	0.8043	0.0000
X_1Y_0	0.9401	0.0008	0.0032	0.0007	0.0080	0.0080	0.0064
X_1Y_1	0.9401	0.0008	0.0032	0.0007	0.8043	0.8043	0.8229
X_1Y_2	0.9401	0.0008	0.0032	0.0007	0.8043	0.8043	0.0000
X_2Y_0	0.9401	0.7737	0.0032	0.0007	0.0080	0.0080	0.0064
X_2Y_1	0.9401	0.7737	0.0032	0.0007	0.8043	0.8043	0.8229
X_2Y_2	0.9401	0.7737	0.0032	0.0007	0.8043	0.8043	0.0000
X_3Y_0	0.9401	0.7737	0.7373	0.0007	0.0080	0.0080	0.0064
X_3Y_1	0.9401	0.7737	0.7373	0.0007	0.8043	0.8043	0.8229
X_3Y_2	0.9401	0.7737	0.7373	0.0007	0.8043	0.8043	0.0000
X_4Y_0	0.9401	0.7737	0.7373	1.0000	0.0080	0.0080	0.0064
X_4Y_1	0.9401	0.7737	0.7373	1.0000	0.8043	0.8043	0.8229
X_4Y_2	0.9401	0.7737	0.7373	1.0000	0.8043	0.8043	0.0000
							Yes Sex No Condom (see *Table 3–2*)

Moreover, the patterns of big ρ parameters (see *Table 3–1*, *Table 3–2*, *Table 3–3* and *Table 3–4*) were consistent with the conceptualization of the two stage sequences. For example, we expected the no use and no sex group would have high probabilities to answer "no" to all items, and this was supported by big ρ parameter estimates (see the first row in *Table 3–1* and *Table 3–3*). Another example is that we expected the no use and safe sex group would have high probabilities to answer "no" to the four items measuring substance use and low probabilities to answer "no" to the two items measuring sexual behavior and one item measuring condom use, but would have a high probability to answer "yes" to always use condoms, and this was also supported. In sum, both adolescent substance use and sexual behavior can be conceptualized as a latent stage sequence with the specified latent statues (please refer back to Chapter 1 or Appendix A for detailed notation about latent status membership X and Y if needed).

Table 3–2: Big ρ parameter estimates for Model A—answer "No" to condom use (*Table 3–1* continued)

X_0Y_0	X_0Y_1	X_0Y_2	X_1Y_0	X_1Y_1	X_1Y_2	X_2Y_0	X_2Y_1	X_2Y_2	X_3Y_0	X_3Y_1	X_3Y_2	X_4Y_0	X_4Y_1	X_4Y_2
0.008	0.005	0.808	0.008	0.005	0.808	0.008	0.005	0.808	0.008	0.005	0.808	0.008	0.005	0.808

Table 3–3: Big ρ parameter estimates for Model B

Status	LifSx1	LifSx2	Con-dom	LifA	RectA	LifM	RectM
	Answer No		No Sex	Answer No			
X_0Y_0	0.9920	0.9920	0.9857	0.8576	0.9992	0.9968	0.9993
X_0Y_1	0.9920	0.9920	0.9857	0.0599	0.9992	0.9968	0.9993
X_0Y_2	0.9920	0.9920	0.9857	0.0599	0.2263	0.9968	0.9993
X_0Y_3	0.9920	0.9920	0.9857	0.0599	0.2263	0.2627	0.9993
X_0Y_4	0.9920	0.9920	0.9857	0.0599	0.2263	0.2627	0.0001
X_1Y_0	0.1957	0.1957	0.1723	0.8576	0.9992	0.9968	0.9993
X_1Y_1	0.1957	0.1957	0.1723	0.0599	0.9992	0.9968	0.9993
X_1Y_2	0.1957	0.1957	0.1723	0.0599	0.2263	0.9968	0.9993
X_1Y_3	0.1957	0.1957	0.1723	0.0599	0.2263	0.2627	0.9993
X_1Y_4	0.1957	0.1957	0.1723	0.0599	0.2263	0.2627	0.0001
X_2Y_0	0.1957	0.1957	0.1919	0.8576	0.9992	0.9968	0.9993
X_2Y_1	0.1957	0.1957	0.1919	0.0599	0.9992	0.9968	0.9993
X_2Y_2	0.1957	0.1957	0.1919	0.0599	0.2263	0.9968	0.9993
X_2Y_3	0.1957	0.1957	0.1919	0.0599	0.2263	0.2627	0.9993
X_2Y_4	0.1957	0.1957	0.1919	0.0599	0.2263	0.2627	0.0001
	Answer Yes		Yes Sex Use Condom	Answer Yes			
X_0Y_0	0.0080	0.008	0.0064	0.1424	0.0008	0.0032	0.0007
X_0Y_1	0.0080	0.008	0.0064	0.9401	0.0008	0.0032	0.0007
X_0Y_2	0.0080	0.008	0.0064	0.9401	0.7737	0.0032	0.0007
X_0Y_3	0.0080	0.008	0.0064	0.9401	0.7737	0.7373	0.0007
X_0Y_4	0.0080	0.008	0.0064	0.9401	0.7737	0.7373	0.9999
X_1Y_0	0.8043	0.8043	0.8229	0.1424	0.0008	0.0032	0.0007
X_1Y_1	0.8043	0.8043	0.8229	0.9401	0.0008	0.0032	0.0007
X_1Y_2	0.8043	0.8043	0.8229	0.9401	0.7737	0.0032	0.0007
X_1Y_3	0.8043	0.8043	0.8229	0.9401	0.7737	0.7373	0.0007
X_1Y_4	0.8043	0.8043	0.8229	0.9401	0.7737	0.7373	0.9999
X_2Y_0	0.8043	0.8043	0.0000	0.1424	0.0008	0.0032	0.0007
X_2Y_1	0.8043	0.8043	0.0000	0.9401	0.0008	0.0032	0.0007
X_2Y_2	0.8043	0.8043	0.0000	0.9401	0.7737	0.0032	0.0007
X_2Y_3	0.8043	0.8043	0.0000	0.9401	0.7737	0.7373	0.0007
X_2Y_4	0.8043	0.8043	0.0000	0.9401	0.7737	0.7373	0.9999
			Yes Sex No Condom (see *Table 3–4*)				

Table 3–4: Big ρ parameter estimates for Model B—answer No to condom use (*Table 3–3* continued)

X_0Y_0	X_0Y_1	X_0Y_2	X_0Y_3	X_0Y_4	X_1Y_0	X_1Y_1	X_1Y_2	X_1Y_3	X_1Y_4	X_2Y_0	X_2Y_1	X_2Y_2	X_2Y_3	X_2Y_4
0.008	0.008	0.008	0.008	0.008	0.005	0.005	0.005	0.005	0.005	0.808	0.808	0.808	0.808	0.808

3.2 Hypotheses 1 and 2

Table 3–5 presents the G-Squared and the degrees of freedom for null Model A-1, Model A-2, Model A-3, Model A-4 and Model A-5, which represented the null hypotheses corresponding to hypothesis 1 and 2, using Model A as the alternative model to express the hypotheses being tested. From *Table 3–5*, it is pretty clear that the five models with equality constraints fit significantly worse than Model A, which had no constraints imposed on conditional probability estimates.

Table 3–5: Model fit index for the null-hypothesis models to be tested against hypothesis 1 and 2

Hypothesis	H_a	H_0 for Hypothesis 1	H_0 for Hypothesis 2 (the first half)		H_0 for Hypothesis 2 (the second half)	
Model	Model A	Model A-1	Model A-2	Model A-3	Model A-4	Model A-5
G-Squared Avrg	3646.8	4211.6	3710.1	3660.8	3661.9	3697.1
G-Squared-Diff		564.8	63.3	14.0	15.1	50.3
df	36729	36737	36737	36733	36733	36733
Df-Diff		8	8	4	4	4

3.2.1 Hypothesis 1

Recall that Model A-1 was the null-hypothesis model to be tested against hypothesis 1—**An adolescent's stage of substance use predicts his/her concurrent stage of sexual behavior**. The freely estimated P($Y|X$) parameter estimates in Model A are presented in *Table 3–6*. The estimates for the same parameters with equality constraints in null Model A-1 are presented in *Table 3–7*.

Table 3–6: The P($Y|X$) parameter estimates in Model A

P(Y|X)	Y_0	Y_1	Y_2
X_0	0.9348	0.0316	0.0336
X_1	0.9378	0.0171	0.0451
X_2	0.7034	0.1612	0.1354
X_3	0.2328	0.4489	0.3182
X_4	0.0646	0.4220	0.5134

Table 3–7: The P($Y|X$) parameter estimates in Model A-1

P(Y|X)	Y_0	Y_1	Y_2
X_0	0.8241	0.0876	0.0883
X_1	0.8241	0.0876	0.0883
X_2	0.8241	0.0876	0.0883
X_3	0.8241	0.0876	0.0883
X_4	0.8241	0.0876	0.0883

Compare the results in the two tables one to one, the increased 564.8 G-Squared with only eight degrees of freedom for Model A-1 was caused simply by forcing the five parameters in each of the three columns to be equal. Thus, Model A is better than Model A-1, meaning that the null hypothesis should be rejected, and the probabilities of being in latent status membership of Y (recall that it was the sexual behavior sequence) should be allowed to be different depending on latent status membership of X (recall that it was the substance use sequence).

To aid the interpretation of the results, *Figure 3–1 is* shown below based on the results given by *Table 3–6*. There we see that the percentages of the subjects in different stages of sexual behavior in grade eight vary considerably depending on where they were in the substance use sequence in the same grade. In short, the results were consistent with hypothesis 1.

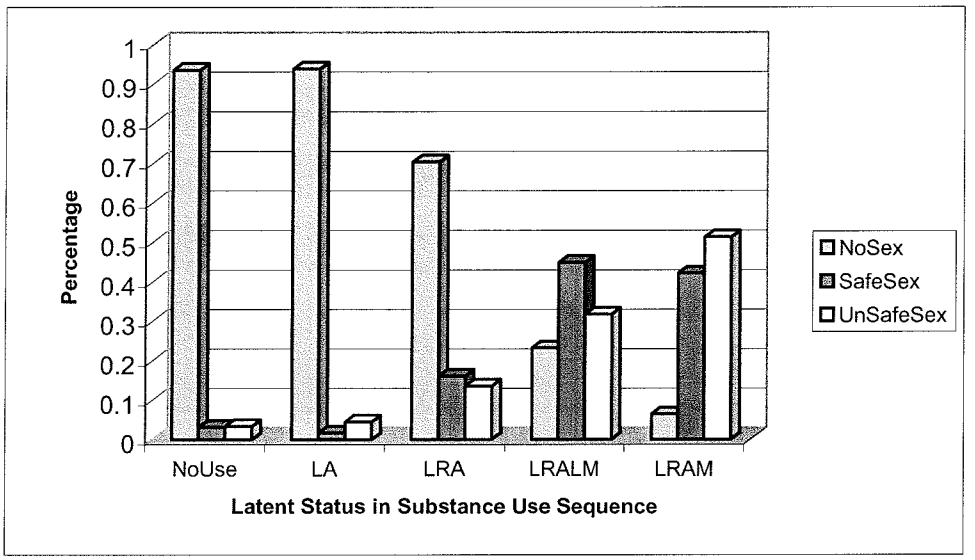

Figure 3–1: Model A P($Y|X$) estimates–showing that the probabilities in stages of the sexual behavior sequence vary considerably with the levels in substance use sequence

3.2.2 Hypothesis 2

Recall that Model A-2, Model A-3, Model A-4, and Model A-5 were the null-hypothesis models to be tested against hypothesis 2—**An adolescent's change in his/her stage of substance use predicts change in his/her stage of sexual behavior, especially when there is advancement into marijuana use.** *Table 3–5* shows that these models with equality constraints all fit significantly worse than Model A, and these null-hypothesis models should be rejected.

The freely estimated P($Y'|X_0Y_0X'$) parameter estimates in Model A are presented in *Table 3–8*. These may be compared with the estimates for the same parameters with

equality constraints in Model A-2 presented in *Table 3–9.* The increase in G-Squared of 63.3 with only eight degrees of freedom for Model A-2 was caused by forcing the five parameters in each of the three columns to be equal. Thus, Model A is a better choice than Model A-2 in terms of model fit, suggesting that the null hypothesis be rejected and that the probabilities of being in various latent statuses of Y should be allowed to differ depending on latent status membership in both sequences in grade eight and latent status membership of X (i.e., the substance use sequence) in grade nine. Recall that X_0Y_0 represents latent membership in no use and no sex in grade eight, so actually, $P(Y'|\, X_0Y_0X')$ parameters give a way to look at the relation between initiation of substance use and subjects' virginity status, and it appears that change in stages of substance use predicts change in subjects' virginity status. Specifically, it appears that the initiation of substance use put the adolescents at a higher risk for onset of sexual behavior. For example, the risk for onset of sexual intercourse was four times more likely for those who initiated alcohol use than for those who remained in no use from grade eight to grade nine. It is even more

Table 3–8: The $P(Y'|\, X_0Y_0X')$ parameter estimates in Model A

$P(Y'\|X_0Y_0X')$	Y_0	Y_1	Y_2
X_0	0.9683	0.0236	0.0081
X_1	0.8707	0.0689	0.0604
X_2	0.9199	0.0801	0.0000
X_3	0.1033	0.5618	0.3349
X_4	0.0005	0.3512	0.6483

Table 3–9: The $P(Y'|X_0Y_1X')$ parameter estimates in Model A-2

$P(Y'\|X_0Y_0X')$	Y_0	Y_1	Y_2
X_0	0.9379	0.0405	0.0216
X_1	0.9379	0.0405	0.0216
X_2	0.9379	0.0405	0.0216
X_3	0.9379	0.0405	0.0216
X_4	0.9379	0.0405	0.0216

striking that the risk for onset of sexual intercourse was twenty-eight times more likely for those who initiated not only alcohol but also marijuana than for those who still remained no use from grade eight to grade nine.

To test the hypothesis that change in stages of substance use predicts change in virginity status, the fit of null Model A-3 was also compared to that of Model A. Results of the freely estimated $P(Y'|X_0Y_1X')$ parameter estimates in Model A are presented in *Table 3–10*. The estimates for the same parameters with equality constraints in Model A-3 are presented in *Table 3–11*. The zeros in the first column in both tables indicate that the parameters were fixed to a value of zero, being consistent with the single-headed arrow between no sex and safe sex or unsafe sex in the sexual behavior sequence (i.e., all backward movements from any sex to no sex were restricted). The $P(Y'|X_0Y_1X')$ parameter estimates tell that, for those who already experienced sex (with condoms) in grade eight, how change in levels of substance use from grade eight to grade nine associates with change in sexual behavior. The significantly decreased fit of Model A-3 suggested that the null hypothesis be

Table 3–10: The $P(Y'|X_0Y_1X')$ parameter estimates in Model A

$P(Y'\|X_0Y_1X')$	Y_0	Y_1	Y_2
X_0	0	0.4652	0.5348
X_1	0	1.0000	0.0000
X_2	0	0.0592	0.9408
X_3	0	0.8466	0.1534
X_4	0	0.4896	0.4104

Table 3–11: The $P(Y'|X_0Y_1X')$ parameter estimates in Model A-3

$P(Y'\|X_0Y_1X')$	Y_0	Y_1	Y_2
X_0	0	0.3883	0.6117
X_1	0	0.3883	0.6117
X_2	0	0.3883	0.6117
X_3	0	0.3883	0.6117
X_4	0	0.3883	0.6117

rejected and that P($Y'|X_0Y_1X'$) parameters should not be forced equal. Therefore, by testing Model A-2 and Model A-3 against Model A, results suggested that not only P($Y'|X_0Y_0X'$) parameters, but also P($Y'|X_0Y_1X'$) parameters should be freely estimated, and that change in stages of substance use predicts change in stages of sexual behavior.

Model A-4 and Model A-5 were designed null-hypothesis models to test the second part of hypothesis 2—advancement into marijuana use put adolescents at a higher risk for initiation of sexual intercourse (please refer back to *Table 2-8* and *Table 2-9* for the equality constraints to test this hypothesis). As predicted, although the two models had the same degrees of freedom, imposing equality constraints on P($Y'|X_0Y_0X_0'$), P($Y'|X_0Y_0X_3'$), and P($Y'|X_0Y_0X_4'$) (i.e., no use, lifetime marijuana, and lifetime and recent marijuana use in grade nine) caused Model A-5 to fit much worse than Model A-4, where equality constraints were imposed on P($Y'|X_0Y_0X_0'$), P($Y'|X_0Y_0X_1'$), and P($Y'|X_0Y_0X_2'$) (i.e., no use, lifetime alcohol, and lifetime and recent alcohol use in grade nine). Recall that from Model A, the risk for onset of sexual intercourse was twenty-eight times greater for those who initiated not only alcohol but also marijuana than for those who remained in no use from grade eight to grade nine, whereas the risk was four times greater for those who initiated only alcohol from grade eight to grade nine. Thus, the worse fit of Model A-5 indicated that advancement into marijuana use put adolescents at a higher risk for initiation of sexual behavior.

3.3 Hypothesis 3 and 4

Table 3–12 presents the G-Squared and the degrees of freedom for Model B-1, Model B-2, and Model B-3, using Model B as the alternative-hypothesis model. The results show that the four models with equality constraints fit significantly worse than Model B, which had no constraints imposed on conditional probability estimates.

Table 3–12: Model fit index for the null-hypothesis models to be tested against hypothesis 3 and 4

Hypothesis	H_a	H_0 for hypothesis 3	H_0 for hypothesis 4	
Model	Model B	Model B-1	Model B-2	Model B-3
G-Squared Avrg	3646.8	4211.9	3710.1	3679.5
G-Squared-Diff		565.1	63.3	32.7
df	36729	36737	36737	36735
Df-Diff		8	8	6

3.3.1 Hypothesis 3

Model B-1 was the null-hypothesis model designed to be tested against hypothesis 3—**An adolescent's stage of sexual behavior predicts his/her concurrent stage of substance use**. The freely estimated P($Y|X$) parameter estimates in Model B are presented in *Table 3–13*. The estimates for the same parameters with equality constraints in Model B-1 are presented in *Table 3–14*.

Table 3–13: The P($Y|X$) parameter estimates in Model B

P($Y|X$)	Y_0	Y_1	Y_2	Y_3	Y_4
X_0	0.5362	0.2601	0.1905	0.0100	0.0031
X_1	0.1722	0.0438	0.4070	0.1841	0.1929
X_2	0.1788	0.1186	0.3398	0.1285	0.2343

Table 3–14: The P($Y|X$) parameter estimates in Model B-1

P($Y|X$)	Y_0	Y_1	Y_2	Y_3	Y_4
X_0	0.4727	0.2286	0.2232	0.0359	0.0397
X_1	0.4727	0.2286	0.2232	0.0359	0.0397
X_2	0.4727	0.2286	0.2232	0.0359	0.0397

Compare the results in the two tables, the increased 565.1 G-Squared with only eight degrees of freedom for Model B-1 resulted from forcing the three parameters in each of the five columns to be equal. So, Model B is more reasonable than Model B-1, namely, the null hypothesis should be rejected and the probabilities of latent status membership of Y (i.e., the substance use sequence) should be allowed to be different depending on latent status membership of X (i.e., the sexual behavior sequence).

Figure 3–2 is shown below based on the results given by *Table 3–13*. The figure shows that the percentages of the subjects in different stages of substance use in grade eight vary considerably depending on their virginity statuses in the same grade. Therefore, the results were not inconsistent with hypothesis 3.

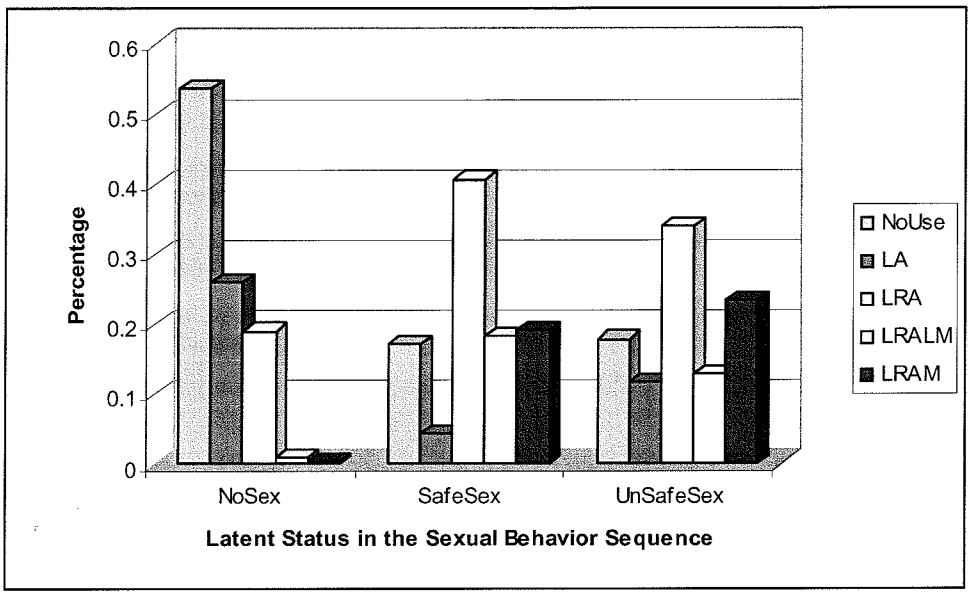

Figure 3–2: Model B P($Y|X$) estimates–showing that the percentages in stages of the substance use sequence vary considerably with the levels in the sexual behavior sequence

3.3.2 Hypothesis 4

Model B-2, and Model B-3 were the null-hypothesis models that intended to be tested against hypothesis 4—**An adolescent's change in his/her stage of sexual behavior predicts change in his/her stage of substance use**. *Table 3–12* suggested that the two models with equality constraints fit significantly worse than Model B. So Model B was a better choice to describe the data, indicating that hypothesis 4 appeared to be supported. *Figure 3–3*, and *Figure 3–4* were based on the results from Model B, where freely estimated P($Y'|\ X_0Y_0X'$), and P($Y'|\ X_0Y_1X'$) parameter estimates were obtained

Figure 3–3 depicts that change from virginity to non-virginity predicts onset of substance use for the subjects from grade eight to grade nine. Further, moving from no

sex to unsafe sex from grade eight to grade nine predicts onset of alcohol and marijuana use.

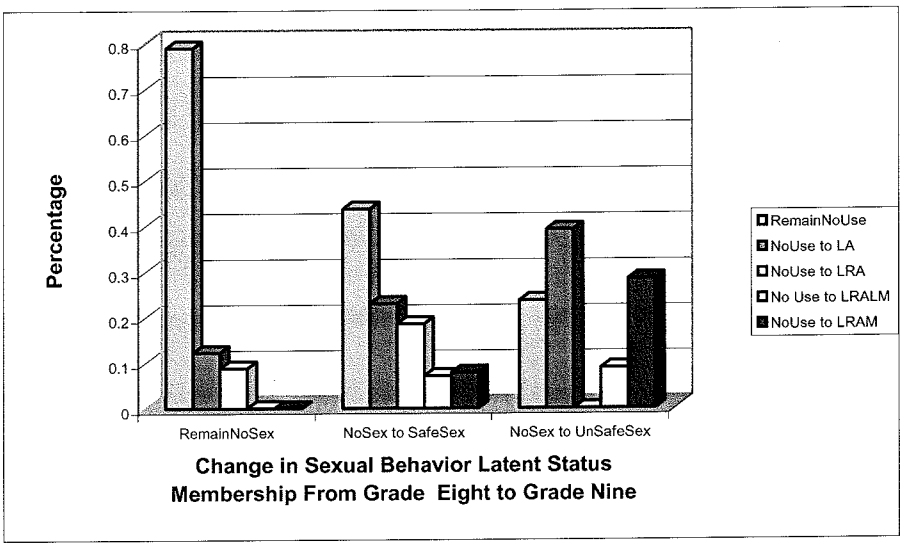

Figure 3–3: Model B-2 P($Y'|X_0Y_0X'$) estimates–showing the relations between change in virginity status and onset of the substance use

Figure 3–4 suggested the relations between change in virginity status and change in stages of the substance use sequence for those who already tried alcohol in grade eight. It appeared that moving to unsafe sex related to both more recent alcohol and marijuana use, whereas advancement to safe sex related to more recent alcohol use not marijuana.

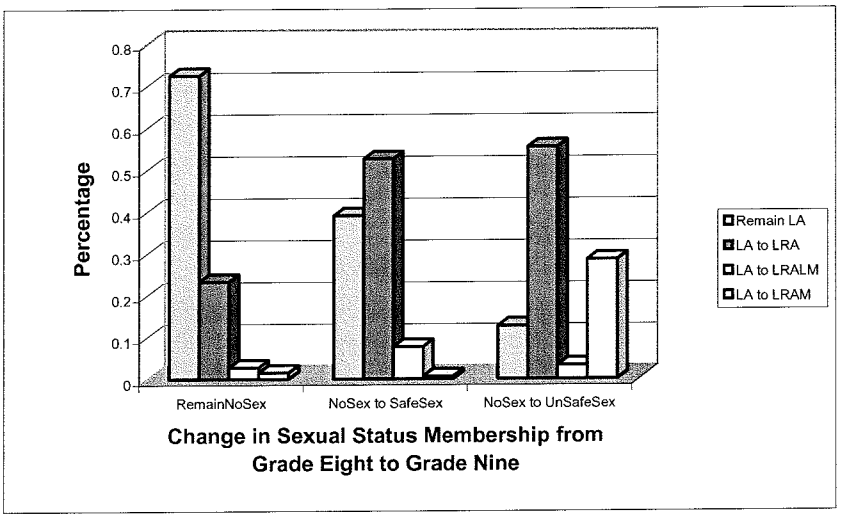

Figure 3–4: Model B-3 $P(Y' | X_0 Y_1 X')$ estimates–showing the relations between change in virginity status and change in substance use

Chapter 4

Discussion and Conclusions

The results showed that both Model A and Model B had a good fit and all the null models with equality constraints on sets of parameters fit significantly worse than Model A and Model B. Therefore, Model A and Model B turned out to be two proper models to describe the relations between substance use and sexual behavior among the subjects in the HFL data set.

Model A and Model B are two stage sequential models, where substance use and sexual behavior were conceptualized as two sequences involving qualitatively different stages or levels, called latent statuses. The involved stages were generally consistent with other studies in the similar areas, particularly for the substance use sequence (e.g., Kandel & Faust, 1975; Kandel, 1989; Kandel, Yamaguchi & Chen, 1992; Kandel & Yamaguchi, 1993; Graham, Collins, Wugalter, Chung, & Hansen, 1991; Collins, Graham, Long & Hansen, 1994; Collins, Graham, Rousculp, & Hansen, 1997; Collins, Hyatt, & Graham, 1998). Few studies conceptualize sexual behavior as a sequence as depicted in the present study. The present study was an effort to consider sexual behavior and substance use jointly in a stage-sequential scenario.

Further, using the HFL data set, the present study shows how to use TS-LTA, an extension of LTA, to model transitions in adolescent substance use and sexual behavior simultaneously. Although there have been many well-known methods such as

regression, SEM, FA, and ANOVA/MANOVA contributing to scientific understanding of the relations between adolescent substance use and sexual behavior, the two-sequence technique presents a different way to look at the same issue. This method has unique features such as allowing latent categorical variables to change over time and modeling changing relations between two changing categorical variables in an asymmetric fashion.

TS-LTA also has several strengths when being compared with other methods dealing with categorical variables, particularly, a contingency table, latent class analysis (LCA; Goodman, 1974; Clogg & Goodman, 1984), and LTA. A contingency table is based on observed or manifest categorical variables, simply cross-tabulating all levels of all items. The contingency table of seven-item two-time repeated measures in this study has $2^{12}*3^2=36864$ possible cells or response patterns, and it is apparently huge. Actually it is impossible to draw the table, and the usual way to present it is using a long vector of observed response patterns. The contingency table only provides frequencies or unconditional probabilities of observed response patterns from raw data. There are several drawbacks associate with such a large contingency table: 1. it is almost impossible to summarize the information provided by data at a theoretical level, 2. it is impossible to conduct hypothesis testing, 3. there is no way to quantify measurement error, and 4. it is impossible to address developmental or longitudinal questions. Therefore, except that one can get some ideas about which observed response patterns are large and which are small, it is rather non-informative to just look at such a large contingency table.

LCA can handle the first three problems that cannot be dealt with by a contingency table. Based on empirical and/or substantive reasons, LCA can impose theoretical structures on a contingency table (similar to SEM and FA) to make the data (i.e., observed response patterns) more interpretable and meaningful, and competing hypothesis regarding possible number of latent classes can be tested (refer to Lanza, Flaherty, & Collins (in press) for details). Further, using multiple observed variables as manifest indicators, LCA can quantify the amount of measurement error and adjust it so that results will be more precise. However, LCA cannot handle the last problem—change and stability or longitudinal development of latent classes. The development of LTA solves this problem, and LTA can handle all the four problems regarding a contingency table. Using LTA, one can not only classify latent classes, but also allow latent classes to change over time (called latent statuses). Both latent classes and latent statuses have been evident in LTA models (e.g., Collins, Graham, Rousculp, & Hansen, 1997; Hyatt & Collins, 2000; Tang, Lanza, & Collins, 2001).

Further, TS-LTA is an extension of LTA, and it can deal with the change and stability of two stage-sequences in stead of only one stage-sequence in LTA. A particular strength of this method compared to LTA is its convenience and flexibility in terms of testing hypotheses as shown in this study. Actually, any two parameters in a two-sequence model, based on substantive reasons, can be tested to see whether they are different or not by comparing two models—one has equality constraints imposed on the two parameters of interest, one does not. Thus, the hypothesis testing can be even more detailed and specific than those that have been conducted in the present study. Such

detailed information cannot be provided by correlation coefficients or LTA parameter estimates. In short, using TS-LTA in this study, the complicated large contingency table with 36864 possible cells or response patterns were not only examined by conducting the four hypotheses testing, but also depicted nicely and meaningfully with the two parsimonious TS-LTA models, namely, Model A and Model B.

In Model A and Model B, measurement errors were pooled across gender and one could argue that this may not be appropriate because females and males may have different probabilities to give an answer to the same question conditional on the same latent status. For example, females maybe more hesitate to report sexual intercourse than males do, so measurement errors may be higher for females than males. Results showed, overall, the measurement was pretty strong in both Model A and Model B, but the two ρ's that had values either below 0.8 at the high end or above 0.2 at the low end might indicate both genders did not necessarily share similar ρ's. This empirical question can be explored in future studies, where gender difference is more of study interest. In the present study, the pooled measurement errors across gender make results easier to be interpreted.

The interesting topic in this study in terms of theoretical aspects was to probe into the relations between substance use and sexual behavior using the HFL data set. In general, the study suggested that substance use predicted concurrent virginity status for adolescents, and onset of substance use predicted onset of sexual intercourse. On the other hand, virginity status predicted concurrent substance use behavior and becoming a non-virgin put adolescents at a higher risk for initiation of substance use. The results

were consistent with the findings from other studies using different methods and samples; for example, studies that suggested substance use and sexual behavior are risk factors for each other using regression (e.g., Allen, Leadbreater & Aber, 1994; Mensch & Kandel, 1992; Rodney, Mupier & O'Neal, 1997; Smith, Udry & Morris, 1985; Capaldi, Crosby & Stoolmiller, 1996; Ensminger, 1990; Rosenbaum & Kandel, 1990; Whitbeck, Yoder, Hoyt, & Conger, 1999), SEM (e.g., Aseltine, 1995; Dishion, Capaldi, & Yoerger, 1999), and FA that shows substance use and sexual behavior cluster together (e.g., Donovan & Jessor, 1985; Donovan, Jessor, & Costa, 1988; Farrell, Danish, & Howard, 1992).

The study indicates the existence of reciprocal effects or symmetric effects of the two sequences for some adolescents, specifically, the substance use sequence predicts the sexual behavior sequence, and the reverse is also true for this particular data set. In other words, both sequences can function as a significant predictor sequence. The results did not appear to show asymmetric effects, where sequence A predicts B but B does not predicts A. However, predicting may be stronger for one direction than for the other even when there are symmetric effects. The ability to detect such "asymmetric effects" with TS-LTA may depend on sample size and enough variability of parameter estimates, and how to demonstrate such asymmetric effects remained to be explored in future studies. Just for Model A and Model B, we notice that both models have the same G-squared with the same degrees of freedom, although the parameter estimates differ from each other in different ways in the two models.

Moreover, with the presence of symmetric effects of the two sequences, it still remains questionable if substance use causes onset of sexual behavior, onset of sexual

intercourse causes substance use, both, or neither. The TS-LTA is intended to disentangle this question, but other things' going on may cast doubt on any causal conclusions. For example, age could be a common cause to link the two behaviors. Generally, substance use increases as age increases during adolescence, showing an age advancing effect (Kandel & Logan, 1984; Graham, Collins, Wugalter, Chung, and Hansen, 1991; Rodney, Mupier, & O'Neal, 1997). A similar age advancing effect for sexual behavior has also been evident in studies (Miller & Moore, 1990; Elliot & Morse, 1989; Capaldi, Crosby, & Stoolmiller, 1996). Although the age range for the subjects in the present study was not very big (nearly 98% were 13-14 years old at grade eight and 14-15 years old at grade nine), one could argue that maturation status varies from individual to individual and it may be a common cause of both substance use and sexual behavior. For example, pubertal development has been found associated with the initiation of substance use (e.g., Hyatt and Collins, 1999) and sexual intercourse (e.g., Smith, Udry, & Morris, 1985; Capaldi, Crosby, & Stoolmiller, 1996). Similarly, researchers (Graber, Petersen & Brooks-Gunn, 1996) argue that "entry into adolescence is marked by the physical changes of puberty" (pp23), and the biological changes can influence developmental trajectories in other domains (e.g., psychosocial, behavioral).

In addition, some researchers (e.g., Donovan & Jessor, 1985; Donovan, Jessor, and Costa, 1988; Farrell, Danish, & Howard, 1992; Costa, Jessor, Donovan, & Fortenberry, 1995) that argue for a general single underlying pattern of proneness/vulnerability to deviance or unconventionality may raise questions regarding why substance use and sexual behavior relate to each other, and unconventionality may

be considered as another common cause. Still some researchers argue that the more risk factors encountered, the higher probability of an adolescent's advancing in his/her level of substance use and sexual experience status (Hawkins, Catalano, & Miller, 1992; Small & Luster, 1994; Sternberg, 1999). Risk factors can be various and they may be people or environments that adolescents interact with on a regular basis such as parents and the family (e.g., Chassin, Rogosch, and Barrera, 1991; Chassin, Curran, Hussong, & Colder, 1996; Hyatt & Collins, 2000; Kendel, 1996), peers (e.g., Donohew, Clayton, Skinner, and Colon, 1999; Aseltine, 1995; Brown, Mory & Kinney, 1994; Kindermann, McCollam & Gibson, 1996), schools, neighborhoods and media (e.g., Allison, Crawford, Leone, Trickett, Perez-Febles, Burton, & Le Blanc, 1999; Aneshensel and Sucoff, 1996; Arnett, 1995; Ward & Rivadeneyra, 1999; Brown, 2000). Given its observational design, this study cannot rule out these possibilities.

Because the data were from survey questionnaires, there is no way to manipulate the independent variable, which is essential to draw any causal conclusion. Causal like conclusions based on survey data, which are basically a correlation type of data, are always questionable. Methods themselves can never solve the problem. For social science phenomena, it is almost impossible to conduct well-controlled experimental studies in order to draw a conclusion with highly internal validity (actually, external validity may be limited by achieving such internal validity). However, based on substantive reasons, one still can develop causal models to speculate relations among variables, and check if they fit data. A model with a good fit such as Model A and Model B in the study can help to understand and address certain questions.

Although it may be somewhat disappointing by not really providing causal conclusions in the end, the TS-LTA is still a good way to depict two changing categorical variables. Compared to a correlation coefficient, this method gives more detailed information as shown in the tables and figures in Chapter 3. From the results, the following conclusions were tentatively drawn: 1. An adolescent's stage of substance use predicts his/her concurrent stage of sexual behavior. 2. An adolescent's change in his/her stage of substance use predicts change in his/her stage of sexual behavior, especially when there is advancement into marijuana use. 3. An adolescent's stage of sexual behavior predicts his/her concurrent stage of substance use. 4. An adolescent's change in his/her stage of sexual behavior predicts change in his/her stage of substance use.

These tentative conclusions may be informative for prevention and intervention research. They provides some information about who, when and what should be the foci of prevention and intervention programs for adolescent problem behaviors. The data were from grade eight and grade nine students, who were in early and middle adolescence, and substance use and sexual behavior were seen intertwined in complicated ways. Thus, prevention and intervention programs aiming at reducing substance use and sexual behavior may need to be administered no later than grade eight and better to have multi-facet components to deal with both problem behaviors. It seems that the older the age range during adolescence, the more intensive is treatment. For those who already involved in and also recently used illicit substance uses such as marijuana, treatment may also need to be more intensive. In addition, using marijuana can put adolescents at higher risk of experiencing other problems (e.g., initiation of sex and experience unsafe sex), so

policy makers may need to consider imposing stringent penalties on providing illicit substances for adolescents.

There are several limitations regarding the present study in terms of illustration of uses of TS-LTA. One is that it did not show the latent class component (e.g., gender, race, and/or others), which can be incorporated within a TS-LTA framework just like in LTA. Another limitation is that only two repeated measures were examined, but actually, TS-LTA can handle more than two measurement occasions. These features can be shown in future studies.

BIBLIOGRAPHY

Allen, J. P., Leadbeater, B. J., & Aber, J. (1994). The development of problem behavior syndromes in at risk adolescents. Development and Psychopathology, 6, 323-342.

Allison, K. W., Crawford, I., Leone, P. E., Trickett, E., Perez-Febles, A. Burton, L. M., & Le Blanc, R. (1999). Adolescent substance use: Preliminary examinations of school and neighborhood context. American Journal of Community Psychology, 27(2), 111-141.

Aneshensel, C. S., & Sucoff, C. A. (1996). The neighborhood context of adolescent mental health. Journal of Health and Social Behavior, 37(4), 293-310.

Arnett, J. J. (1995). Adolescents' use of the media for self-socialization. Journal of Youth and Adolescence, 24, 519-533.

Aseltine, R. H. Jr. (1995). A reconsideration of parental and peer influences on adolescent deviance. Journal of Health and social Behavior, 26(2), 103-121.

Bingham, C. R., & Crockett, L. J. (1996). Longitudinal adjustment patterns of boys and girls experiencing early, middle, and late sexual intercourse. Developmental Psychology, 32(4), 647-658.

Brooks-Gunn, J., & Furstenberg, F. F. Jr. (1989). Adolescent sexual behavior. American Psychologist, 44(2), 249-257.

Brown, J. D. (2000). Adolescents' sexual media diets. Journal of Adolescent Health, 27(2), 35-40.

Brown, B. B., Mory, M. S., & Kinney, D. (1994). Casting adolescent crowds in relational perspective: Caricature, channel, and context. In R. Montemayer, G. Adams, & T. P. Gullotta (Eds.), Personal relationships during adolescence, (pp.123-167). Newbury Park, CA: Sage

Capaldi, D. M., Crosby, L., & Stoolmiller, M. (1996). Predicting the timing of first sexual intercourse for at-risk adolescent males. Child Development, 67, 344-359.

Chassin, L., Curran, P. J., Hussong, A. M., & Colder, C. R.(1996). The relation of parent alcoholism to adolescent substance use: A longitudinal follow-up study. Journal of Abnormal Psychology, 105(1), 70-80.

Chassin, L., Rogosch, F., & Barrera, M. (1991). Substance use and symptomatology among adolescent children of alcoholics. Journal of Abnormal Psychology, 100(4), 449-463.

Clogg, C. C., & Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional contingency tables. Journal of the American Statistical Association, 79, 762-771.

Collins, L. M., Graham, J. W., Long, J. D. & Hansen, W. B. (1994). Crossvalidation of latent class models of early substance use onset. Multivariate Behavioral Research, 29, 165-183.

Collins, L. M., Graham, J. W., Rousculp, S. S., & Hansen, W. B. (1997). Caffeine use and the beginning of the substance use onset process: An illustration of latent transition analysis. In K. Bryant, M. Windle, & S. West (Eds.), The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research (pp79-99). Washington, DC: American Psychological Association.

Collins, L. M., Hyatt, S. L. & Graham, J. W. (1998). Latent transition analysis as a way of testing models of stage-sequential change in longitudinal data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), Modeling Longitudinal and Multiple-Group Data: Practical Issues, Applied Approaches, and Specific Examples (pp147-161). Hillsdale, NJ: Erlbaum.

Collins, L. M. & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. Multivariate Behavioral Research, 27 (1), 131-157.

Costa, F., Jessor, R., Donovan, J., & Fortenberry, J. (1995). Early initiation of sexual intercourse: The influence of psychosocial unconventionality. Journal of Research on Adolescence, 5, 93-121.

Dempster, A.P., Laird, N. M. & Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of The Royal Statistical Society Series B, 39, 1-38.

DiClemente, R. J., Hansen, W. B., & Ponton, L., E. (Ed.) (1996). Handbook of Adolescent Health Risk Behavior. New York: Plenum Press.

Dishion, T. J.; Capaldi, D. M.; Yoerger, K.(1999). Middle childhood antecedents to progressions in male adolescent substance use: An ecological analysis of risk and protection. Journal of Adolescent Research, 14(2), 175-205.

Donohew, L., Clayton, R. R., Skinner, W. F., & Colon, S. (1999). Peer networks and sensation seeking: Some implications for primary socialization theory. Substance Use and Misuse, 34(7), 1013-1023.

Donovan, J. E., & Jessor, R. (1985). Structure of problem behavior in adolescence and young adulthood. Journal of Consulting and Clinical Psychology, 53(6), 890-904.

Donovan, J. E., Jessor, R., & Costa, F. M. (1988). Syndrome of problem behavior in adolescence: A replication. Journal of Consulting and Clinical Psychology, 56(5), 762-765.

Elliott, D. S., & Morse, B. J. (1989). Delinquency and drug use at risk factor in teenage sexual activity. Youth and Society, 21(1), 32-60.

Ensminger, M. E. (1990). Sexual activity and problem behaviors among Black, urban adolescents. Child Development, 61, 2032-2046.

Farrell, A.D., Danish, S. J., & Howard, C. W. (1992). Relationship between drug use and other problem behaviors in urban adolescents. Journal of Consulting and Clinical Psychology, 60, 705-712.

Farrell, A. D., & White, K. S. (1998). Peer influences and drug use among urban adolescents: Family structure and parent-adolescent relationship as protective factors. Journal of Consulting and Clinical Psychology, 66(2), 248-258.

Flaherty, B. P., & Collins, L. M. (1999). Modeling transitions in two stage-sequences simultaneously. Technical report series #99-33. The Methodology Center, The Pennsylvania State University.

Frankel, M. R., McWilliams, H. A. & Spencer, B. D. (1983). NLSY technical sampling report & addendum (1994). University of Chicago.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 61, 215-231.

Graber, J. A., Petersen, A. C., & Brooks-Gunn, J. (1996). Pubertal processes: Methods, measures, and models. In J. A. Graber, J. Brooks-Gunn, & A. C. Peterson (Eds.), Transitions Through Adolescence: Interpersonal Domains and Context (pp23-53). Mahwah, NJ: Lawrence Earbaum Associates, Pubilishers.

Graham, J. W., Collins, L. M., Wugalter, S.E., Chung, N. K., & Hansen, W. B. (1991). Modeling transitions in latent stage-sequential processes: A substance use prevention sample. Journal of Consulting and Clinical Psychology, 59, 48-57.

Graham, J. W., Rohrbach, L., Hanson, W. B., Flay, B. R., & Johnson, C. A. (1989). Convergent and discriminant validity for assessment of skill in resisting a role play alcohol offer. Behavioral Assessment, 11, 353-379.

Graves, K. L., & Leigh, B. C. (1995). The relationships of substance use to sexual activity among young adults in the United States. Family Planning Perspectives, 27(1)18-22+33.

Harter, S. (1990). Self and identity development. In S. S. Feldman & G. R. Elliot (Eds.), At the threshold: The developing adolescent, (pp. 352-387). Cambridge, Mass: Harvard University Press.

Hyatt, S. L., & Collins, L. M. (1999). Pubertal Timing and the Onset of Substance Use: Detailed Results Based on Data Augmentation. Technical report series #99-35. The Methodology Center, The Pennsylvania State University.

Hyatt, S. L., & Collins, L. M. (2000). Using latent transition analysis to examine the relationship between parental permissiveness and the onset of substance use. In J. Rose, L. Chassin, C. Presson, & S. Sherman (Eds.), Multivariate Applications in Substance Use Research. Hillsdale, NJ: Erlbaum, pp. 259-288.

Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. Psychological Bulletin, 112, 64-105.

Jessor, R. (1998). New perspectives on adolescent risk behavior. In R. Jessor (Ed.), New Perspectives on Adolescent Risk Behavior. New York: Cambridge University Press.

Jessor, R., & Jessor, S. (1977). Problem Behavior and Psychological Development: A Longitudinal Study of Youth. New York: Academic Press.

Kendel, D. B. (1996). The parental and peer contexts of adolescent deviance: An algebra of interpersonal influences. Journal of Drug Issues, 26(2), 289-315.

Kandel, D. B. (1989). Issues of sequencing of adolescent drug use and other problem behaviors. Drugs and Society, 3, 55-76. New York, NY: Hawthorne Press.

Kandel, D. B., & Faust, R. (1975). Sequences and patterns of adolescent drug use. Archives of General Psychiatry, 32, 923-932.

Kandel, D. B., & Yamaguchi, K. (1993). From beer to crack: Developmental patterns of drug involvement. American Journal of Public Health, 83, 851-855.

Kandel, D. B., Yamaguchi, K. & Chen, K. (1992). Stages of progression in drug involvment from adolescence to adulthood: further evidence for the gateway theory. Journal of Studies on Alcohol, 447-457.

Kandel, D., B., & Logan, J. A. (1984). Patterns of drug use from adolescence of young adulthood: I. Periods of risk for initiation, continued use and discontinuation. American Journal of Public Health, 74(7), 660-666.

Kindermann, T. A., McCollam, T. L., & Gibson, E. Jr. (1996). Peer networks and students' classroom engagement during childhood and adolescence. In J. Juvonen

& K. R. Wenzel (Eds.), Social Motivation: Understanding children's school adjustment, (pp.279-312). New York: Cambridge University Press.

Little, R. J. A. and Rudin, D. B. (1987). Statistical Analysis with Missing Data. J. Wiley & Sons, New York.

McGinnis, J. M., & Foege, W. H. (1993). Actual causes of death in the United States. Journal of American Medical Association, 270(18), 2207-2212.

Mensch, B. & Kandel, D. B. (1992). Drug use as a risk factor for premarital teen pregnancy and abortion in a national sample of young white women. Demography, 29(3), 409-429.

Miller, B., & Moore, K. (1990). Adolescent sexual behavior, pregnancy, and parenting: Research through the 1980s. Journal of Marriage and the Family, 52, 1025-1044.

Moberg, D. P., Piper, D. L., Wu, J., & Serlin, R. C. (1993). When total randomization is not possible: Nested randomized assignment. Evaluation Review, 17, 271-291.

Mott, F. L., Fondell, M. M., Hu, P. N. Kowaleski-Jones, L., & Menaghan E. G. (1996). The determinants of first sex by age 14 in a high-risk adolescent population. Family Planning Perspectives, 28(1), 13-18.

Mott, F. L., & Haurin, R. J. (1988). Linkages between sexuality and alcohol and drug use among American adolescents. Family Planning Perspectives, 20(3), 128-136.

Piper, D. L., Moberg, D. P., & King, M. J. (2000). The healthy for life project: Behavioral outcomes. The Journal of Primary Prevention, 21(1), 47-73.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., Tabor, J., Beuhring, T., Sieving, R. E., Shew, M., Ireland, M., Bearinger, L. H., Udry, J. R. (1997). Protecting adolescents from harm: Findings from the National Longitudinal Study of Adolescent Health. Journal of the American Medical Association, 278, 823-832.

Rodney, H. E., Mupier, R., and O'Neal, S. (1997). African-American youth in public housing showing low alcohol and drug use. Journal of Child and Adolescent Substance Abuse, 6(4), 55-73.

Rosenbaum, E. & Kandel, D. (1990). Early onset of adolescent sexual behavior and drug involvement. Journal of Marriage and the Family, 52, 783-798.

Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys. New York, NJ: J. Wiley & Sons.

Schafer, J. L. (1997). Analysis of incomplete multivariate data. London: Chapman & Hall.

Small, S., & Luster, T. (1994). Adolescent sexual activity: An ecological, risk-factor approach. Journal of Marriage and the Family, 56, 181-192.

Smith, E., Udry, J., & Morris, N. (1985). Pubertal development and friends: A biosocial explanation of adolescent sexual behavior. Journal of Health and Social Behavior, 26, 183-192.

Sternberg, L. D. (1999). Adolescence. The McGraw-Hill Companies, Inc. Stewart M. (1994). High school students who use drugs are more likely than nonusers to engage in risky sexual behavior. Family Planning Perspectives, 26(6), 279-280.

Stewart M. (1994). High school students who use drugs are more likely than nonusers to engage in risky sexual behavior. Family Planning Perspectives, 26(6), 279-280.

Tang, Z., Lanza, S., & Collins, L. M. (2001). Modeling adolescent substance use using latent transition analysis: The healthy for life data set. Technical report series #01-44. The Methodology Center, The Pennsylvania State University.

Ward, L. M., & Rivadeneyra, R. (1999). Contributions of entertainment television to adolescents' sexual attitudes and expectations: The role of viewing versus viewer involvement. The Journal of Sex Research, 36, 237-249.

Weissberg, R. P., Barton, H. A., & Shriver, T. P. (1997). The social –competence promotion program for young adolescents. In George W. Albee & Thomas P. Gullitta (Eds.), Primary prevention works. Thousand Oaks: Sage publications.

Whitbeck, L. B., Yoder, K. A., Hoyt, D. R., & Conger, R. D. (1999). Early adolescent sexual activity: A developmental study. Journal of Marriage and the Family, 61, 934-946.

Zabin, L. S., Hardy, J. B., Smith, E. A., Hersch, M. B. (1986). Substance use and its relation to sexual activity among inner-city adolescents. Society for Adolescent Medicine, 7, 320-331.

Appendix A

Notation for the models

Table A–1 Notation for Model A

Sequence	*Latent Status*	*Math Notation*	*Meaning*
Predictor	N	X_0	No Use
	LA	X_1	Lifetime Alcohol
	LRA	X_2	Lifetime and Recent Alcohol
	LRALM	X_3	Lifetime and Recent Alcohol and Lifetime Marijuana
	LRAM	X_4	Lifetime and Recent Alcohol and Marijuana
Dependent	No Sex	Y_0	No Sex
	Safe Sex	Y_1	Safe Sex
	Unsafe Sex	Y_2	Unsafe Sex

Table A–2 Notation for Model B

Sequence	*Latent Status*	*Math Notation*	*Meaning*
Predictor	No Sex	X_0	No Sex
	Safe Sex	X_1	Safe Sex
	Unsafe Sex	X_2	Unsafe Sex
Dependent	N	Y_0	No Use
	LA	Y_1	Lifetime Alcohol
	LRA	Y_2	Lifetime and Recent Alcohol
	LRALM	Y_3	Lifetime and Recent Alcohol and Lifetime Marijuana
	LRAM	Y_4	Lifetime and Recent Alcohol and Marijuana