Multivariate Behavioral Research, *32* (3), 275-318

Latent Variable Modeling of Longitudinal and Multilevel Substance Use Data

Terry E. Duncan, Susan C. Duncan, Anthony Alpert, and Hyman Hops
Oregon Research Institute, Eugene, Oregon

Mike Stoolmiller
Oregon Social Learning Center, Eugene, Oregon

Bengt Muthén
Graduate School of Education & Information Studies
University of California, Los Angeles

This article demonstrates the use of a general model for latent variable growth analysis which takes into account cluster sampling. Multilevel Latent Growth Modeling (MLGM) was used to analyze longitudinal and multilevel data for adolescent and parent substance use measured at four annual time points. An associative LGM model was tested for alcohol, marijuana, and cigarette use with a sample of 435 families. Hypotheses concerning the shape of the growth curve and the extent of individual differences in the common trajectory over time were tested. The effects of marital and family status and socio-economic status on family levels of substance use were also examined. Findings are discussed in terms of family-level substance use and similarities in developmental trajectories across substances, and the impact of contextual factors on family levels of substance use and development.

Over the last three decades, we have witnessed a gradual increase in the complexity of theoretical models that attempt to explain problem behavior, such as alcohol use and abuse in children and adolescents. There has been marked movement away from an emphasis on person variables and the focus on nonreciprocal parent influences to a greater examination of person-environment interactions (e.g., Bronfenbrenner, 1977).

Bell (1968) was one of the earliest proponents of a familial model that emphasized the reciprocal exchanges within families that could exert

This research was supported by Grant No. DA 03706 from the National Institute on Drug Abuse. Preparation of this manuscript was supported in part by Grant No. DA 09548-01 from the National Institute on Drug Abuse.

Requests for reprints and all correspondence should be addressed to Terry E. Duncan, Oregon Research Institute, 1715 Franklin Blvd., Eugene, OR, 97403, Tel.: (541) 484-2123, FAX: (541) 484-1108. Electronic mail may be sent via Internet to terryd@ori.org.

influence on all family members. Other influential perspectives include that of family systems (Minuchin, 1974) and social interactionist (Patterson, 1984, 1986) theorists, both of whom maintain that all family members are part of an interactive, interdependent network in which the behavior of each individual, or subsystem, modifies that of other individuals, or subsystems, in the family. Contributing to this process may be the social modeling of problem behaviors such as substance use, by other family members, as purported by social learning (e.g., Akers & Cochran, 1985) and problem-behavior (Jessor & Jessor, 1977) theories.

Findings from the literature suggest that there are remarkable similarities in adjustment outcomes among family members in general (e.g., Ary, Tildesley, Hops, & Andrews, 1993; Duncan, Duncan, & Hops, 1996; Gfroerer, 1987; Melby, Conger, Conger, & Lorenz, 1993; Needle et al., 1986). Thus, it is assumed that families characterized by poor adolescent adjustment and problem behavior define a somewhat homogeneous social environment shared by all of its members, parents and siblings alike.

Interdependence of Family Data

Despite the assumption of a somewhat homogeneous, shared environment of the family, psychological research until recently has had few tools to accommodate the interdependence of family data. The absence of hierarchical methodologies for handling interdependence of families' data led to researchers' reliance on analyses utilizing data for a "target individual", even after collecting data on all members of the social unit under investigation (e.g., Biernat & Wortman, 1991).

This article considers data collected in such a hierarchical fashion. Multivariate modeling of such data are most frequently done as if the data were obtained as a simple random sample from a single population. Hence, the standard assumption of independent and identically distributed observations is made. However, analyzing the data as a simple random sample ignores the potential interdependence within families or clusters and can lead to inflated test statistics for estimated parameters and overall model fit. A hierarchical approach avoids these distortions. New analytic techniques that are more suited to the hierarchical data structure have recently emerged under the labels of hierarchical, or multilevel models (see e.g. Aitkin & Longford, 1986; Burstein, 1980; de Leeuw & Kreft, 1986; Duncan, Duncan, Hops, & Stoolmiller, 1995; Goldstein 1986; Longford, 1987; Raudenbush & Bryk, 1988; Schmidt & Wisenbaker, 1986). Muthén and Satorra (1989) point out that such models take into account the correlated observations and observations from heterogeneous populations with varying parameter values, gathered in hierarchical data. While appropriate analysis techniques of this kind are now available for standard regressions and analysis of variance situations, Muthén and Satorra highlighted the lack of techniques for covariance structure models, such as factor analysis, path analysis, and structural equation modeling (SEM).

Muthén (1989) discussed the relationships of multilevel SEM to conventional SEM and pointed out the possibility of using conventional SEM software for multilevel structural equation modeling. In the special case of balanced data (i.e., all group sizes are equal), the multilevel SEM model can be estimated by formulating it as a conventional two sample, multiple group model using within group and between group covariance matrices. However, if the data are unbalanced (i.e., there are many groups and most groups are of different size), maximum likelihood (ML) estimation with hierarchical data is computationally heavy. Moreover, conventional software may not work at all if the software requires positive definite input matrices (between group covariance matrices will not be positive definite if the number of variables is greater than the number of observations) and the input specifications can be very tedious. Muthén (1991, 1994) has proposed an ad hoc estimator, using limited information, that is simpler to compute than maximum likelihood while still using the SEM framework. Muthén (1994) shows that this simpler estimator is consistent, is identical to the ML estimator for balanced data and gives results close to those of ML for data that is not too badly unbalanced. This suggests that with many group sizes little may be gained by the extra effort of ML computation. Taken together, these developments make possible the construction, estimation, and testing of a variety of complex models utilizing hierarchically structured data without developing new specialized software.

Interrelationships in the Development of Substance Use

Evidence for the interrelationships among adolescent problem behaviors is substantial. Positive correlations have been found among the following behaviors: antisocial behavior, cigarette smoking, alcohol use, use of marijuana and other illicit drugs, precocious sexual activity and risky sexual behaviors, dangerous driving, poor school performance, and general risk taking (e.g., Bachman, Johnston, & O'Malley, 1981; Biglan, et al., 1990; Brennan, 1979; Donovan & Jessor, 1985; Donovan, Jessor, & Costa, 1988; Farrell, Danish, & Howard, 1992; Hawkins, Lishner, Catalano, & Howard, 1986; Jessor, 1987a, 1987b; Jessor & Jessor, 1977; Loeber & Dishion, 1983; Osgood, Johnston, O'Malley, & Bachman, 1988; Zabin, 1984). Multivariate analyses have shown that a single common factor can account for the relationships among these behaviors (Donovan & Jessor, 1985; Donovan et al., 1988; Farrell et al., 1992; Osgood et al., 1988), and where gender differences have been investigated, these interrelationships have been found to hold for both males and females (Donovan & Jessor, 1985; Farrell et al., 1992).

Despite this body of evidence that various problem behaviors are intercorrelated, most studies examining the covariance among the various problem behaviors have been static, that is, developed primarily from measures taken at only one point in time. However, with increased interest in the development of substance use, greater emphasis has been placed on the time dimension and the development of dynamic models pertaining to both intra- and inter-individual development of substance use and its etiology during adolescence. The literature shows consistently that adolescent substance use increases with age followed by a gradual decrease in use during the middle 20s continuing through adulthood (e.g., Johnston, O'Malley, & Bachman, 1993). As older adolescents become more active in the peer group, spending disproportionately more time with them (Berndt, 1982; Montemayor & Brownlee, 1987) and under their influence, increased substance use is not unexpected.

Recent developments in statistical techniques expand the opportunity to examine trends and individual differences in substance use, and to explore the effects of the social context on these developmental trends. One methodology which provides a means of modeling individual differences in growth curves is termed a latent growth model or LGM. Recent studies by Duncan and Duncan (1994) and Duncan, Tildesley, Duncan, and Hops (1995) demonstrate the utility of LGM for determining trends in adolescent alcohol, cigarette, and marijuana use across five- and four-year periods, respectively. Using an associative model which incorporated all three substances simultaneously, these data provided initial support for similarities in the development of these substances during adolescence. Other applications of LGM may be found in Duncan and Duncan (1995), Duncan, Duncan, and Hops, (1994), Duncan and McAuley (1993), Duncan and Stoolmiller (1993), McArdle (1988), McArdle and Epstein (1987), McArdle and Hamagami (1991), and Stoolmiller, Duncan, Bank, and Patterson (1993). Different approaches to growth modeling can be found in Bryk and Raudenbush (1987), Hops, Duncan, Duncan, and Stoolmiller (1996), and Willet, Ayoub, and Robinson (1991).

In the present study, we extend previous research in substance use development (e.g., Duncan et al., 1994) using latent growth modeling. Using new methodology developed in Muthén (1997) for multilevel latent growth modeling (MLGM) we define an associative growth curve model for alcohol, marijuana, and cigarette use over a four-year period. Using multilevel analyses, developmental changes in alcohol, marijuana and cigarette use are examined among adolescents and parents from 435 families. It was expected that there would be sufficient homogeneity in the development of substance use among family members to adequately define a two-factor unspecified latent growth model (LGM). It was also expected that the level of heterogeneity among families in their development of substance use would be sufficient to conduct a multilevel latent growth model (MLGM) analysis. Three family-level variables, marital status (single or two-parent), family type (step- and foster families or biologically related families) and socio-economic status, were included to examine how much of the heterogeneity among families could be explained by these variables. Based on research indicating that alcohol use tends to start in early adolescence, progressing to more regular use throughout middle to late adolescence followed by a gradual decrease in use during the middle 20s (following a possible curvilinear rather than linear trajectory) and that alcohol use is higher among boys compared to girls (e.g., Byram & Fly, 1984; Curran, Hartford, & Muthén 1996; Duncan et al., 1994; Robbins & Martin, 1993), gender, and linear and quadratic effects for age were controlled for in the model by their inclusion as predictors of the within-level variance in alcohol use variables.

Method

Subjects

Data in this study are part of a longitudinal study on the predictors and consequences of substance use among adolescents (Duncan et al., 1994; Duncan, Duncan, Hops, et al., 1995). Participants were from two northwestern urban areas of the United States, with populations of approximately 50,000 and 120,000. Recruitment was through advertisements in local newspapers and fliers posted in community centers and various street locations. Families contained at least one adolescent between the ages of 11 and 15 who was designated as the target. In order to participate in the study, siblings had to be at least 11 years of age. For the purpose of the present study, data from 435 families (435 target adolescents, 203 siblings, and 566 parents (168 fathers and 398 mothers) were included in the analyses. Adolescents (targets and siblings) comprised 312 males and 326 females, with a mean age at *T*1 of 13.69 years (SD = 1.95). Forty-eight percent of the adolescents were from single-parent families, while the remainder resided with parents who were married or living in a committed relationship.

A majority of the participants were white (90%) and came from skilled or professional families (Hollingshead, 1975). Comparison of our adolescent sample's substance use with national norms, as estimated by the National Institute on Drug Abuse National Household Survey on Drug Abuse (NIDA, 1989), tends to be higher, but is representative of the region estimated from the Oregon Employment Division Research and Statistics (1989) report on eighth and eleventh graders.

Procedures

Targets, siblings, and parents separately completed a series of self-report questionnaires in separate rooms in our laboratory. The questionnaires were designed to assess substance use and selected psychosocial characteristics. Families were paid $35 for their participation. To ensure and increase the likelihood of valid reporting of substance use by parents and adolescents, confidentiality was stressed at each assessment and a certificate of confidentiality was obtained from the National Institute on Drug Abuse that precluded the subpoena of subjects' data.

Measures

Target and Sibling Substance Use

Self-report measures of target and sibling alcohol, cigarette, and marijuana use were each assessed via a 5-point scale which was developed from items measuring status and frequency of use of each drug. The five levels on the scales represent: (0) life time abstainers; (1) 6-month abstainers; (2) Current use of less than four times a month; (3) Current use of between 4 and 29 times a month; (4) Current use of 30 or more times a month. Although these scales were created from status and frequency information, the assumption is made that the underlying properties are continuous in nature.[1]

Parent Substance Use

To accurately assess the higher level of adult substance use relative to adolescent use, parents reported their average quantity of daily use of

[1] For more information on this scale see Duncan & Duncan (1994), Duncan et al. (1994), and Duncan, Duncan, Hops, et al. (1995).

cigarettes, alcohol, and marijuana. A similar procedure can be found in Newcomb & Bentler (1988). For parents, self-report scales ranging from (0) life time abstainers to (6) daily use, for alcohol, (0) life time abstainers to (4) smoke more than a pack a day, for cigarettes, and (0) life time abstainers to (4) use more than once a day, for marijuana, were utilized. In these data, for equivalence of scaling between parents and adolescents, the parent alcohol scale was divided by 1.5, resulting in a 0-4 scaling.

Contextual Variables

Parental marital status family status, and socio-economic status (SES) were included as predictors of the between-level variation in substance use scores. SES was computed as the average of parental income and education. Parental income was measured with a 16 point scale which assessed a range of annual income from "6,000 dollars and below" to "50,000 dollars or more". To assess educational attainment, parents were asked to report the highest level of education completed. Responses to education were coded on a 7 point scale that ranged from "Grade level 6 or less" to "Graduate level". Items were standardized and then averaged to create the measure of SES. Marital status was coded as follows: married or living in a committed relationship = 1; single = 0. Family status was coded as follows: step or foster families = 0; other = 1.

Within-level Predictors

Age and gender were controlled for in the model by their inclusion as predictors of the within-level variance in substance use variables. Age was a continuous variable representing age at *T*1 and gender was a dichotomous variable where 0 = males and 1 = females. Because of the possibility of a non-linear relationship between age and the various substance use variables, a quadratic term for age was also included as a variable in the various model tests.

Descriptive statistics are presented in Table 1 (next page). As shown in the table, univariate values of skewness and kurtosis are in most cases, minimal, an indication that the assumptions of approximate normality of the observed variables is tenable. Approximate normality justifies the use of normal theory maximum likelihood estimation techniques found in SEM programs such as EQS (Bentler, 1990).

Table 1
Descriptive Statistics

	Mean	Variance	Skewness	Kurtosis
Family Aggregates				
Marijuana *T1*	1.58	.82	1.44	1.26
Marijuana *T2*	1.64	.84	1.26	.78
Marijuana *T3*	1.74	.96	1.20	.75
Marijuana *T4*	1.82	.99	1.04	.37
Cigarettes *T1*	1.97	1.45	1.22	.52
Cigarettes *T2*	2.06	1.47	1.08	.22
Cigarettes *T3*	2.12	1.50	1.05	.17
Cigarettes *T4*	2.27	1.67	.86	-.30
Alcohol *T1*	1.99	1.18	.23	-.60
Alcohol *T2*	2.08	1.23	.20	-.67
Alcohol *T3*	2.14	1.32	.16	-.72
Alcohol *T4*	2.27	1.48	.04	-.76
SES				
Income	13.99	92.20	.33	-1.19
Education	7.85	8.99	.32	-.95
Marital Status (single vs other)	.65	.23	-1.59	-.65
Family Status (step parent vs other)	.13	.55	-1.15	-.21

Results

Latent Growth Models

Latent growth curve models are basically variants of the standard linear structural model; strongly resembling the classic confirmatory factor analysis. However, because they use repeated measures raw-score data, the latent factors are actually interpreted as chronometric common factors representing individual differences over time (McArdle, 1988). Meredith and Tisak (1990) note that repeated measures polynomial analysis of variance (ANOVA) models are actually special cases of latent growth models in which only the factor means are of interest. In contrast, a fully expanded latent growth analysis takes into account both factor means and variances. This combination of the individual and group levels of analysis is unique to the procedure.

Because LGM techniques have been described in detail elsewhere (e.g., Duncan & Duncan, 1994, 1995; Duncan et al., 1994; McArdle, 1988; McArdle & Anderson, 1989; McArdle & Epstein, 1987; McArdle & Hamagami, 1991; Meredith & Tisak, 1990; Stoolmiller, 1994; Willet & Sayer, 1994), only a brief description of LGM will be presented here.

Figure 1 represents a latent growth model appropriate to the present study, in which observed variables measured at four different occasions are represented by two common factors. The first common factor is labeled the intercept, which represents the point at which the growth curve intercepts the vertical axis at the initial time point. The intercept factor represents information in the sample concerning the mean, represented by *Mi*, and variance, represented by *Di*, of the collection of individual intercepts that characterize each individual's growth curve. The second factor has a mean, *Ms*, and variance, *Ds*, and represents the slope or shape of an individual's trajectory over time. The means of the intercept and slope factors represent group growth parameters. The variance of the latent factors reflects the variation of each individual subject around the group mean. The basic LGM therefore, can be considered a random coefficients model. The intercept and slope are allowed to covary, *Ris*, which is represented by the double headed arrow between the two factors.

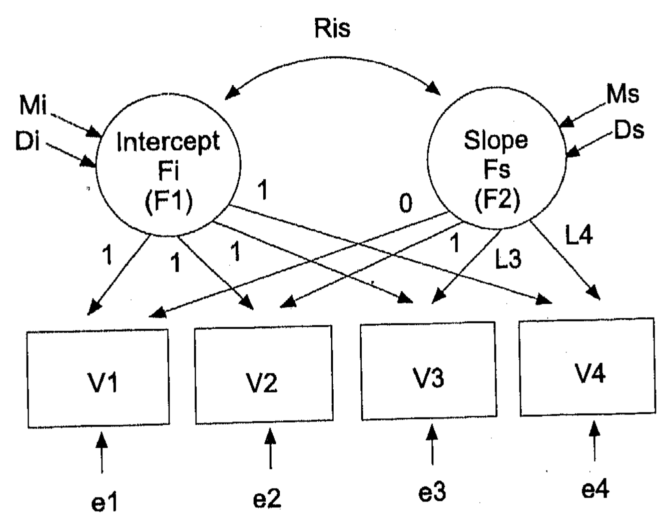


Figure 1
Representation of a Two-factor Unspecified LGM for Four Points in Time

The scaling of the slope can be controlled by the choice of loadings on this factor. However, we chose to fit an unspecified curve, a developmental function that reflects a set of parameters which represent an optimal patterning (freely estimated loadings) over occasions for the changes in substance use scores (Meredith & Tisak, 1990). The estimated parameters within this approach then reflect the developmental function with maximal fit to the data (Rao, 1958; Tucker, 1958). An unspecified growth curve is a "crooked line", a piecewise linear curve. The unspecified model can be quite useful and powerful because it is capable of approximating quadratic and other curvilinear trajectories. The general two-factor latent growth modeling approach has many advantages which recommend its use in the testing and evaluation of developmental models. Using this parameterization, investigators can study predictors of change separately from correlates of initial status. Example applications and more details on the unspecified model can be found in Duncan & Duncan (1994), Duncan et al. (1994) and Meredith and Tisak (1990).

The factor loadings plotted against the observed time metric give a clue as to the form or shape of growth. The following are details concerning the statistical model for the LGM depicted in Figure 1. The mathematical model, with subscripts for individuals suppressed, can be symbolized as follows:

$$Vt = Fi + LtFs + Et, \quad t = 1, ..., 4,$$

where Vt = observed score at time t, Fi = unobserved score for the intercept factor, Fs = unobserved score for the growth rate factor, Et = unobserved composite error at time t, Lt = basis coefficient for time t, in essence, the factor loading relating observed variables Vt to latent growth variables. Observed scores are specified to be a weighted sum of three individual latent variables.

1. Fi is a latent variable representing individual differences in the level of some attribute which is constant for any individual across time.

2. Fs is a latent shape/slope variable representing individual differences in the rate of change of some attribute over time. Like the Fi score, Fs is a constant for any individual across time. The contribution of Fs to Vt, however, changes as a function of the basis coefficient Lt.

3. Et is an unobserved composite error term with a mean of zero and zero correlations with all other variables over time. Such composite errors represent both random errors of measurement and time specific influences for any individual.

The basis term Lt is a mathematical function relating observed variables V to latent growth variables specified in terms of linear departures from an origin, and may be of linear or nonlinear forms. To illustrate a simple linear basis function, let

$$L = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$$

represent the function for the data at times 1 through 4 respectively. The equations for $V1$ through $V4$ are therefore,

$$\begin{aligned} V1 &= Fi + 0Fs + E1, \\ V2 &= Fi + 1Fs + E2, \\ V3 &= Fi + 2Fs + E3, \\ V4 &= Fi + 3Fs + E3, \end{aligned}$$

and to illustrate a concrete example, if $Fi = 2$, $Fs = 1$ and $Et = 0$, $t = 1, \ldots, 4$, then for this individual the model implies the following trajectory:

$$\begin{aligned} V1 &= 2 + 0(1) + 0 = 2 \\ V2 &= 2 + 1(1) + 0 = 3 \\ V3 &= 2 + 2(1) + 0 = 4 \\ V4 &= 2 + 3(1) + 0 = 5, \end{aligned}$$

where $L1 = 0$ at time 1 simply starts the curve at this point by rescaling the intercept factor to represent initial status, $L2 = 1$ at time 2 indicates that from time 1 to time 2 there is one unit of change, $L3 = 2$ indicates that from time 1 to the third time point there are 2 units of change, and $L4 = 3$ indicates that from time 1 to the fourth time point there are 3 units of change. Thus, Lt describes a linear relation of change in terms of linear differences from initial status at time 1.

We should point out that although it is possible to keep adding factors until a satisfactory fit to the data is obtained, latent growth modeling, in general, is more powerful if a small number of factors can adequately describe the data. The question of how many factors it takes to represent a given growth form, or how well a small number of factors approximates any particular type of nonlinear trajectory, is an issue that is beyond the scope of this article. Interested readers should consult Tucker (1958), Tisak and Meredith (1990), or Burchinal and Appelbaum (1991).

Multilevel Latent Growth Modeling (MLGM)

The new methodology in Muthén (1997) involves two generalizations of previous SEM work: (a) SEM growth modeling as generalized to cluster data, and (b) SEM multilevel modeling as generalized to mean structures, where Muthén shows that means will appear in the between structure while within structure means are constrained at zero.

As pointed out in Muthén (1989), MLGM is a complex analysis which needs to follow a sound strategy. In our example, the total covariance matrix, $\mathbf{\Sigma}_T$, is decomposed into two independent components, a between families covariance matrix, $\mathbf{\Sigma}_B$, and a within families covariance matrix, $\mathbf{\Sigma}_W$, or $\mathbf{\Sigma}_T = \mathbf{\Sigma}_B + \mathbf{\Sigma}_W$. Conventional covariance structure analysis that ignores the grouping or clustering assumes that all observations are independent which implies that $\mathbf{\Sigma}_B = 0$. Thus, Muthén (1994) recommended that the actual MLGM be preceded by four important analysis steps: conventional factor analysis of total structure using the sample total covariance matrix, $\mathbf{S}_T$, estimation of between family variation or level of intraclass correlation (ICC), estimation of within family structure, and estimation of between family structure. The analyses in the present study were performed utilizing the EQS (Bentler, 1990) structural equation modeling program, and a Fortran routine (SOURCEBW) developed by Muthén and colleagues (Muthén, 1992, 1994; Nelson & Muthén, 1991) which calculates the necessary within- and between family covariance and mean structures and provides ICC estimates and the ad hoc estimator needed for model estimation. The program code and instructions for the use of this program are given in Nelson and Muthén (1991). (Sample output from SOURCEBW for our example can be retrieved via anonymous ftp from any web browser via the URL: ftp://ftp.ori.org/pub/bw). We note that the standard errors and chi-square tests of model fit are not exact but are reasonable approximations. Consequently, statements about significance and model fit should not be interpreted in exact terms (Muthén, 1997).

Step 1: Conventional Factor Analysis of the Total Structure

This analysis is useful in generating and testing various model ideas. The analysis is, however, incorrect when the data are multilevel due to the correlated observations. The chi-square test of model fit is usually inflated, particularly for data with large intraclass correlations, large family sizes, and highly correlated variables. However, the test of fit may be of practical use by giving a rough sense of fit.

The S_T analysis of the associative two-factor unspecified LGM yields a reasonable fit, $\chi^2(33, N = 1204) = 159.20, p \leq .001$, NNFI = .98, CFI = .99, given the large sample size of 1204. This sample size makes the power of the test high and thus rejection at the 5% level might have reflected trivial deviations from the model. Results indicated an upward trend in the development of adolescent and parent alcohol, cigarette, and marijuana use, consistent with other developmental studies (e.g., Duncan et al., 1994; Duncan, Tildesley et al., 1995). In addition, both initial levels of use and the developmental trajectories among the various substances were significantly interrelated. These findings lend support to other literature (e.g., Duncan, Tildesley et al., 1995; Hansen, Graham, Sobel, Shelton, Flay, & Johnson, 1987; Tildesley, Hops, Ary, & Andrews, 1995) documenting the similarities between adolescent's use of various substances. We note that, despite the adequate fit of the model to the data, this test is incorrect, given that the hierarchical nature of the data is ignored.

Step 2: Estimation of Between Variation

When considering all 435 families in the sample, there are 5 distinct clusters (family) sizes, ranging from 2 to 6. Of interest is whether a multilevel analysis of these data is warranted. This can be accomplished by testing $\Sigma_B = 0$ which can be carried out in an MLGM. A simpler process, however, is to compute the estimated intraclass correlations for each variable. These estimates may be obtained by random effects ANOVA (Winer, Brown, & Michels, 1991) or the SOURCEBW program developed by Muthén (Nelson & Muthén, 1991). The intraclass correlation for a given variable is the ratio of the between family variance to the total variance. Thus, if all intraclass correlations are close to zero, it might not be worthwhile to entertain multilevel analyses. In testing the level of ICC using a random effects ANOVA, a significant effect of the grouping variable in accounting for total variation in the dependent measure would provide indication that a multilevel approach was tenable. However, failure to detect significant family effects in the random effects ANOVA should not preclude conducting an MLGM. The ability of the MLGM to account for measurement error at multiple levels of the hierarchy may lead to different conclusions about between level effects, especially when variable reliability is questionable. A good overview of intraclass correlation estimation is given in Koch (1983). The intraclass correlations, for the four repeated measures were .27, .31, .34, and .32 for marijuana ($V1$, $V2$, $V3$, $V4$); .25, .26, .27, and .29 for cigarettes ($V5$, $V6$, $V7$, $V8$); and .27, .21, .18, and .12 for alcohol use ($V9$, $V10$, $V11$, $V12$). Of interest is the effect gender and the linear and quadratic effects of age will have on the estimates of between variation as these effects will be included as within level predictors in the MLGM. These conditional intraclass correlations, for the four repeated measures were .27, .30, .33, and .30 for marijuana (*V*1, *V*2, *V*3, *V*4); .25, .25, .26, and .28 for cigarettes (*V*5, *V*6, *V*7, *V*8); and .30, .28, .28, and .26 for alcohol use (*V*9, *V*10, *V*11, *V*12), making it reasonable to proceed to Step 3.

Step 3: Estimation of Within Structure $\mathbf{S}_{PW}$

The third step carries out the analysis of the sample pooled-within matrix, $\mathbf{S}_{PW}$. The pooled-within matrix, $\mathbf{S}_{PW}$, was obtained from Muthén's SOURCEBW program. This analysis estimates individual-level parameters only. Muthén (1994) has shown that the analysis gives estimates that are close to the within parameters of an MLGM. The conventional analysis would set the sample size at the total number of observations minus the number of families and either the normal theory GLS or ML estimator. Since the $\mathbf{S}_{PW}$ analysis is not distorted by the between covariation, it is expected to give a better model fit than the $\mathbf{S}_T$ analysis (see also Keesling & Wiley 1974; Muthén 1989) and is, therefore, the preferred way to explore the individual-level variation. Conventional ML analysis gives an adequate fit for the two-factor $\mathbf{S}_{PW}$ model, $\chi^2(33, N = 769) = 84.23, p < .001$, NNFI = .98, CFI = .99. Compared to the test of $\mathbf{S}_T$, $\mathbf{S}_{PW}$ demonstrates a considerable, $\chi^2 = 74.97$, drop in the chi-square test statistic. Muthén (1989) notes that, relative to the conventional analysis of $\mathbf{S}_T$, the $\mathbf{S}_{PW}$ analysis adjusts for differences in family means, whereas in the $\mathbf{S}_T$ analysis, heterogeneity in the means across families increases the total variation, which inflates the reliabilities of the individual measures.

Step 4: Estimation of Between Structure $\mathbf{S}_B$

The fourth step investigates the between structure using $\mathbf{S}_B$, the sample between families covariance matrix. In general, the analysis of the between structure can be the more difficult part of multilevel analysis because it does not concern the customary individual-level data but instead across-family (co)variation. As analyses in Cronbach (1976) and Harnqvist (1978) have shown, the same structure as that seen in the within level cannot automatically be expected as the between components can have a different meaning than the within components. In our particular application, the between family variation concerns between family variation in trajectories of substance use which is conceptually quite similar to individual variation in substance use trajectories. Thus, we expect a similar growth structure for both within and between levels although we allow parameter estimates to differ across the two components.

Some complications arise in exploring the structure of $\mathbf{S}_B$. First, as pointed out by Muthén (1994), $\mathbf{S}_B$ actually estimates $\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B$. For balanced data, c is the common family size. For unbalanced data and a large number of families, c is close to the mean of the family sizes. Thus, one can analyze $\mathbf{S}_B$ to get a notion of the $\mathbf{\Sigma}_B$ structure, but any simple structure expected to hold for $\mathbf{\Sigma}_B$ does not necessarily hold for $\mathbf{S}_B$. Second, $(\mathbf{S}_B - \mathbf{S}_{PW})/c$ is an ML estimate of $\mathbf{\Sigma}_B$ and it might be tempting to analyze this matrix instead of $\mathbf{S}_B$. Unfortunately, the ML estimate of $\mathbf{\Sigma}_B$ is frequently nonpositive definite and may even have negative variance estimates. Muthén (1994) has noted that when it is possible to analyze both $(\mathbf{S}_B - \mathbf{S}_{PW})/c$ and $\mathbf{S}_B$, the results have been similar. Clearly, exploratory analyses of between family structure is an area that needs more development and we are studying the possibility of using graphical techniques. With these caveats in mind, the associative LGM analysis using the $\mathbf{S}_B$ matrix generated from SOURCEBW and a two-factor unspecified model, resulted in a model fit of $\chi^2(33, N = 435) = 95.50$, $p < .001$, NNFI = .97, CFI = .98. The non-normed fit index (NNFI) and comparative fit index (CFI) both reflect fit relatively well at all sample sizes, avoiding underestimation of fit sometimes found in true models with normed fit indices. Estimates greater than .90 in magnitude are generally indicative of a model which adequately describes the relationships observed in the observed data. Reported values, therefore, suggest an adequate fit for the two-factor unspecified between structure. The concepts behind these fit indices are laid out in more detail in Bentler (1990).

The four initial analysis steps suggest an associative MLGM model with a two-factor unspecified latent growth structure for both within and between levels. The MLGM makes use of $\mathbf{S}_{PW}$ and $\mathbf{S}_B$ simultaneously. The specification of this model for marijuana, cigarettes, and alcohol can be illustrated by a diagram such as that in Figure 2 (next page).[2] In line with

[2] The multilevel LGM as presented can be considered a three-level random coefficients model with repeated observations nested within subjects, nested within families. Using notation from Bryk and Raudenbush (1992), a comparable model for estimating the growth of one substance, without predictors at the between and only the basis term specified at the within level, can be denoted as:

Level 1

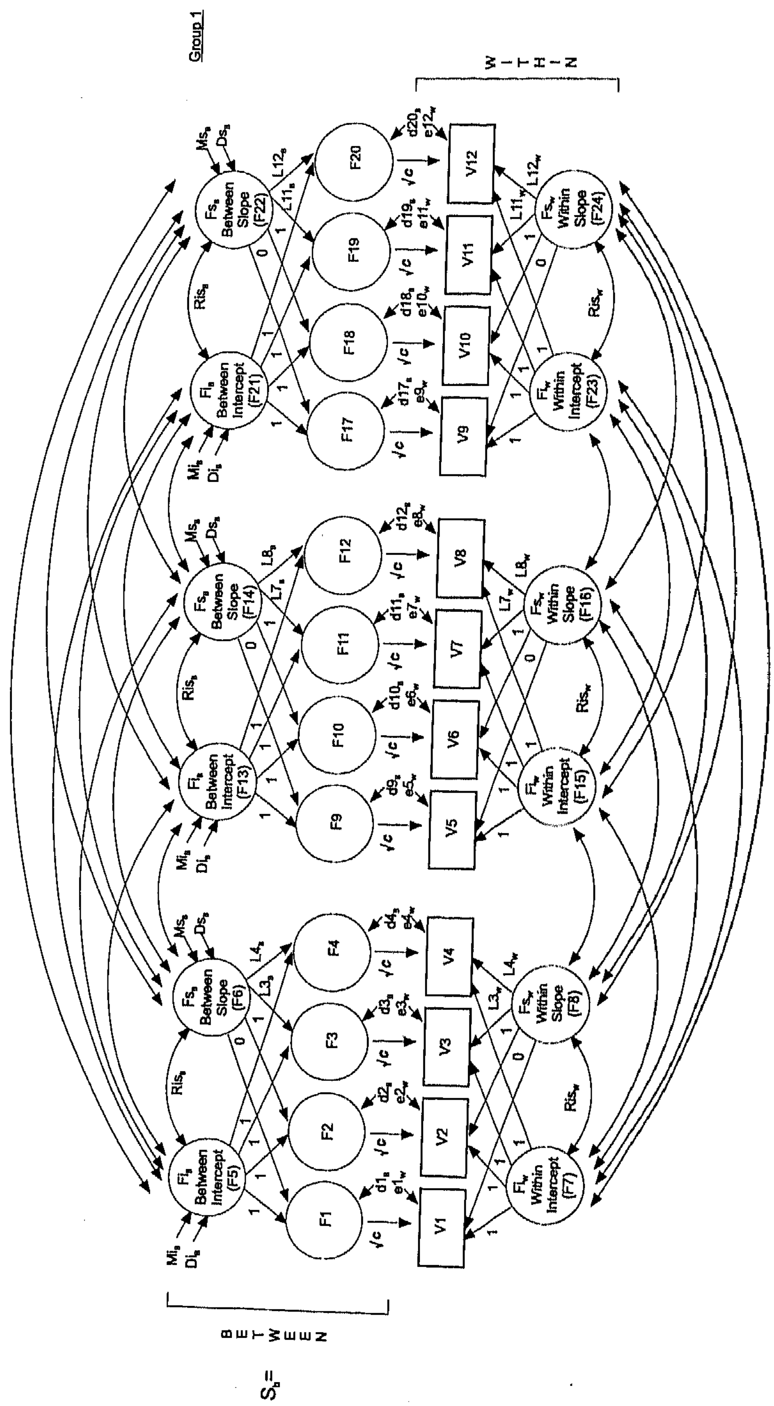
$Y_{tjk} = \pi_{0jk} + \pi_{1jk} + e_{tjk}$, where Y_{ijk} = the observed score at time t for subject j in family k. π_{0jk} = initial status for subject j, π_{1jk} = specified basis term, and e_{ijk} = random error.

Level 2

$\pi_{0jk} = \beta_{00j} + r_{0ij}$ and $\pi_{1jk} = \beta_{10j} + r_{1ij}$, where β_{00j} = mean initial status for family j, β_{10j} = mean growth rate for family j and r is the individual variation from families.

Level 3

$\beta_{00j} = \gamma_{000} + u_{00j}\beta_{10j} = \gamma_{100} + u_{10j}$, where γ_{000} = grand mean intercept, γ_{100} = grand mean growth rate, and u = family variability.


Group 1
WITHIN
BETWEEN
S_b=

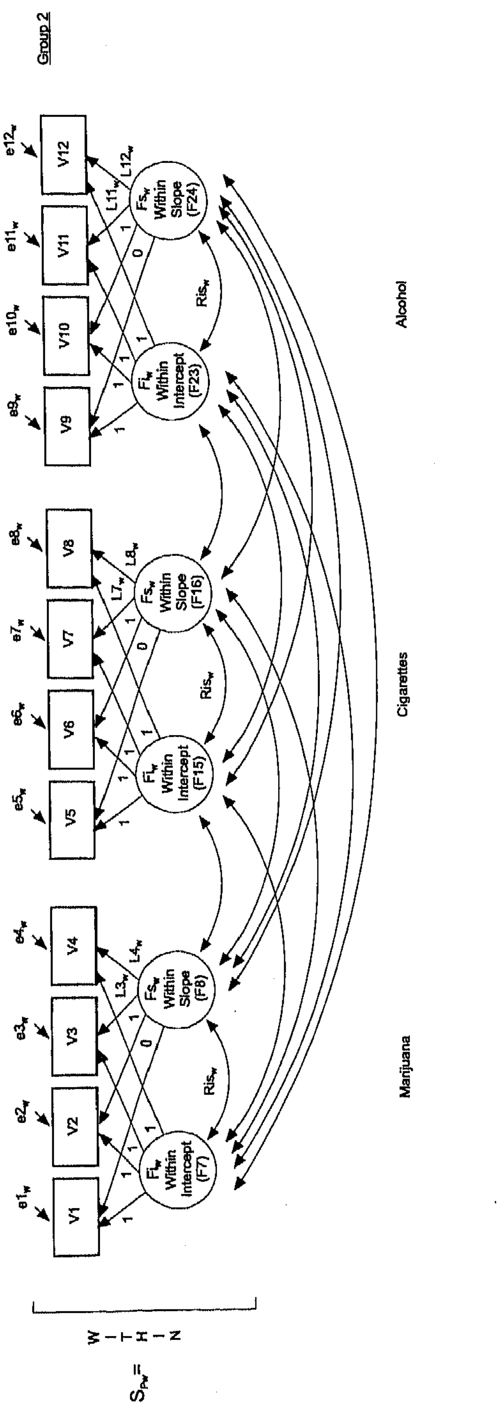


Figure 2
(Divided across pages) Representation of a Multi-level Associative Two-factor Unspecified LGM Model Using Within- and Between-level Components
Note. Within-level predictors of gender and linear and quadratic terms for age, presented in Figure 2, are not shown here for the sake of clarity.

structural modeling conventions, squares denote observed variables and circles denote latent variables. The part of the model below the squares refers to the within structure, while the part above refers to the between structure. The set up for the first group involves both between and within level structure but the set up for the second group involves only within level structure. In the first model only the substance use variables and within family predictors of gender and linear and quadratic terms for age were used. Between family contextual variables of SES and marital status were included as predictors of between family variation in substance use trajectories in a second model.

We will specify the following model for the i^{th} individual in the g^{th} family where G is the total number of families, the number of observations within the g^{th} family is N_g, and the total number of observations is N = summation N_g,

$$
\begin{bmatrix} V1_{gi} \\ V2_{gi} \\ V3_{gi} \\ V4_{gi} \\ V5_{gi} \\ V6_{gi} \\ V7_{gi} \\ V8_{gi} \\ V9_{gi} \\ V10_{gi} \\ V11_{gi} \\ V12_{gi} \end{bmatrix}
=
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & L3_B & 0 & 0 & 0 & 0 \\
1 & L4_B & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & L7_B & 0 & 0 \\
0 & 0 & 1 & L8_B & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & L11_B \\
0 & 0 & 0 & 0 & 1 & L12_B
\end{bmatrix}
\begin{bmatrix} M5i_B \\ M6s_B \\ M13i_B \\ M14s_B \\ M21i_B \\ M22s_B \end{bmatrix}
+
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & L3_B & 0 & 0 & 0 & 0 \\
1 & L4_B & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & L7_B & 0 & 0 \\
0 & 0 & 1 & L8_B & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & L11_B \\
0 & 0 & 0 & 0 & 1 & L12_B
\end{bmatrix}
$$

$$
\begin{bmatrix} F5i_B \\ F6s_B \\ F13i_B \\ F14s_B \\ F21i_B \\ F22s_B \end{bmatrix}
+
\begin{bmatrix} D1_{Bg} \\ D2_{Bg} \\ D3_{Bg} \\ D4_{Bg} \\ D9_{Bg} \\ D10_{Bg} \\ D11_{Bg} \\ D12_{Bg} \\ D17_{Bg} \\ D18_{Bg} \\ D19_{Bg} \\ D20_{Bg} \end{bmatrix}
+
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & L3_W & 0 & 0 & 0 & 0 \\
1 & L4_W & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & L7_W & 0 & 0 \\
0 & 0 & 1 & L8_W & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & L11_W \\
0 & 0 & 0 & 0 & 1 & L12_W
\end{bmatrix}
\begin{bmatrix} F7i_W \\ F8s_W \\ F15i_W \\ F16s_W \\ F23i_W \\ F24s_W \end{bmatrix}
+
\begin{bmatrix} E1_W \\ E2_W \\ E3_W \\ E4_W \\ E5_W \\ E6_W \\ E7_W \\ E8_W \\ E9_W \\ E10_W \\ E11_W \\ E12_W \end{bmatrix}.
$$

The first two terms to the right of the equals sign capture the contribution of the means of the between family slope and intercept factors to the means of the observed variables after multiplication by the between family basis coefficients. The third and fourth terms correspond to between-family variation in the observed V variables due to the between families growth process, and the fifth term is the between family error term. The sixth and seventh terms correspond to within-level variation in the observed V variables due to the within-level growth process, and the eighth term is the within-level error term. Notice that factors Fi and Fs are specified on both the between family level, Fi_B and Fs_B, ($F5$ *and* $F6$ for marijuana; $F13$ *and* $F14$ for cigarettes; and $F21$ *and* $F22$ for alcohol) and the within family level, Fi_W and Fs_W, ($F7$ *and* $F8$ for marijuana; $F15$ *and* $F16$ for cigarettes; and $F23$ *and* $F24$ for alcohol). The basis terms, L_W and L_B, are the coefficients for the influence of the between and within slope factors on the observed V variables, and indicate that non-linear growth can be accommodated for both between and within slopes separately by estimating the 3rd and 4th coefficients, holding the first two values fixed at 0 and 1, respectively, for identification purposes. The influence of the slope factors need not be the same on the between side as it is on the within side.

In terms of actual estimation using conventional SEM software, the model is set up as a two sample, multiple groups problem with $\mathbf{S}_B$ input for the first group and $\mathbf{S}_{PW}$ input for the second group. Modeling is done with $\mathbf{S}_B$ as opposed to the ML estimate of $\mathbf{\Sigma}_B$ for reasons noted above in the discussion of the fourth preliminary step.

Considering the three sample covariance matrices $\mathbf{S}_T$, $\mathbf{S}_{PW}$, and $\mathbf{S}_B$,

$$\mathbf{S}_T = (N - 1)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{N_g} (y_{gi} - \bar{y})(y_{gi} - \bar{y})'$$

$$\mathbf{S}_{PW} = (N - G)^{-1} \sum_{g=1}^{G} \sum_{i=1}^{N_g} (y_{gi} - \bar{y}_g)(y_{gi} - \bar{y}_g)'$$

$$\mathbf{S}_B = (G - 1)^{-1} \sum_{g=1}^{G} N_g(\bar{y}_g - \bar{y})(\bar{y}_g - \bar{y})'.$$

The matrix $\mathbf{S}_T$ is used in conventional covariance structure analysis. In the multilevel case it can be considered a consistent estimator of the total covariance matrix $\mathbf{\Sigma}_W + \mathbf{\Sigma}_B$. Muthén (1994), has demonstrated that the pooled-within matrix $\mathbf{S}_{PW}$ is a consistent and unbiased estimator of $\mathbf{\Sigma}_W$, while the between matrix $\mathbf{S}_B$ is a consistent and unbiased estimator of $\mathbf{\Sigma}_W + C\mathbf{\Sigma}_B$, where C reflects the family size.

$$C = \left[N^2 - \sum_{g=1}^{G} N^2 g \right] [N(G-1)]^{-1}.$$

For unbalanced data, C is close to the mean of the family sizes. We note that the between family matrix $\mathbf{S}_B$ is the covariance matrix of family means weighted by the family size. Therefore, the ML estimate of $\mathbf{\Sigma}_W$ is $\mathbf{S}_{PW}$, while the ML estimate of $\mathbf{\Sigma}_B$ is (Muthén, 1990)

$$C^{-1}(\mathbf{S}_B - \mathbf{S}_{PW}).$$

The fitting function that is minimized is

$$G\{ln|\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B| + trace[(\mathbf{\Sigma}_W + c\mathbf{\Sigma}_B)^{-1}\mathbf{S}_B] - ln|\mathbf{S}_B| - p\} +$$
$$(N - G)[ln|\mathbf{\Sigma}_W| + trace(\mathbf{\Sigma}_W^{-1}\mathbf{S}_{PW}) - ln|\mathbf{S}_{PW}| - p].$$

The top equation corresponds to the first group in the multiple group set up which captures the between level contribution to the total variation and is weighted by G, the number of families. The bottom equation corresponds to the second group and captures the within level contribution to the total variation and is weighted by $N - G$, the total sample size minus the number of families. We note that for balanced data, c is the common family size and the two group approach above is equivalent to full information maximum likelihood estimation.

The model set up in the first group using $\mathbf{S}_B$ as input requires the creation of extra latent variables in order to capture the weighting by the constant c. These extra latent variables are depicted as $F1$ through $F4$ for marijuana, $F9$ through $F12$ for cigarettes, and $F17$ through $F20$ for alcohol in Figure 2, and their contribution to the observed V variables is scaled by fixing the path (loading) to $\sqrt{c}$. The residual variances of $F1$ through $F4$ ($D1$ through $D4$), $F9$ through $F12$ ($D9$ through $D12$), and $F17$ through $F20$ ($D17$ through $D20$), capture the between level composite error variances. The mean structure for the MLGM arises from the four observed variable means for each substance being expressed as functions of the means of the Fi_B and Fs_B between level factors. The means of the within level growth factors are fixed at zero.

The second group in the multiple group set up corresponds to the within variation. The covariance structure of Σ_W is captured by using the same model structure as for the first group, following Figure 1, but fixing all between level coefficients and variance-covariance parameters to zero. Within level composite error variance is captured by residual error variances, $E1_W$ through $E12_W$ for the V variables. Because Σ_W also appears in the covariance structure of the first group, equality restrictions across groups need to be applied for the within parameters. For a complete exposition and computational details concerning the ad hoc estimation procedures see Muthén (1991, 1994). Input specifications for estimation of the MLGM using the EQS (Bentler, 1990) structural equation modeling program can be found in Appendix A.

Model fitting procedures for the associative two-factor unspecified latent growth model fit simultaneously to the data for alcohol, cigarette, and marijuana use, resulted in a chi-square test statistic of $\chi^2(147, N = 1204) = 390.72, p < .001$, NNFI = .98, CFI = .98.

Considering the number of observed variables in the model, 15, and the complexity of the associative MLGM analyses, this model was considered acceptable. Model parameters are presented in Table 2 (next page). As can be seen from the table, all slope means were significant showing evidence of substantial development in the use of all three substances. Basis terms for the three substances were estimated for both between and within-level components. These estimates are shown in Table 3.

The latent variances were also estimated. The variances for intercepts and slopes are provided for all three substances at the between and the within level. These estimates, shown in Table 2 suggest that significant variation existed in individual differences regarding initial status and developmental trajectories for all three substances. For the between level, this indicates that substantial heterogeneity exists across families.

Considering the between to total factor variance ratio, approximately 32%, 27% and 35% of the total variation in marijuana, cigarette, and alcohol intercept scores, respectively, not accounted for by the individual-level predictors could be accounted for by the individuals' family membership. Similarly, approximately 12%, 23% and 22% of the total variation in marijuana, cigarette, and alcohol slope scores, respectively, could be accounted for by the individuals' family membership.

Estimated between-level correlations for initial status and slope, for all three substances, derived from the simultaneous testing of the growth curves for each substance in the associative model, are presented in Table 4. Within-level correlations are presented in Table 5. As expected, levels of initial status for all three substances are highly interrelated. Those

Table 2
Parameter Estimates for the Hypothesized Multilevel LGM

Coefficient	Between-level Effect	Between-level t-value	Within-level Effect	Within-level t-value
Growth Parameters				
Means				
Marijuana Intercept	1.58	49.91		
Marijuana Slope	.05	3.24		
Cigarettes Intercept	1.97	47.26		
Cigarettes Slope	.07	4.57		
Alcohol Intercept	2.00	51.73		
Alcohol Slope	.08	4.21		
Variances				
Marijuana Intercept	.21	7.06	.46	14.84
Marijuana Slope	.006	1.67	.044	3.47
Cigarettes Intercept	.35	6.62	.95	17.39
Cigarettes Slope	.01	2.20	.04	3.549
Alcohol Intercept	.32	7.15	.59	14.34
Alcohol Slope	.01	1.60	.04	3.33
Covariates				
Age → Marijuana Intercept			1.97	8.49
Age → Marijuana Slope			-1.36	-4.33
Age → Cigarettes Intercept			1.12	4.95
Age → Cigarettes Slope			-1.41	-4.59
Age → Alcohol Intercept			.16	.65
Age → Alcohol Slope			-1.96	-5.84
Age Quadratic → Marijuana Intercept			-1.96	-8.47
Age Quadratic → Marijuana Slope			.97	3.25
Age Quadratic → Cigarettes Intercept			-1.01	-4.46
Age Quadratic → Cigarettes Slope			.93	3.28
Age Quadratic → Alcohol Intercept			-.43	-1.77
Age Quadratic → Alcohol Slope			1.35	4.51
Gender → Marijuana Intercept			.10	.28
Gender → Marijuana Slope			-.03	-.76
Gender → Cigarettes Intercept			-.10	-.30
Gender → Cigarettes Slope			-.17	-.42
Gender → Alcohol Intercept			-.09	-2.58
Gender → Alcohol Slope			.06	1.57

Note. t-values greater than 1.96 (1.65 for variances) in magnitude indicate a parameter estimate which is significantly different from zero. Parameter estimates for the effects of the covariates are presented as standardized regression coefficients. All other parameter estimates are presented in unstandardized form.

Table 3
Growth Functions for the Hypothesized Associative Multilevel LGM

	Time					
	1	2	3		4	
			Lt	*t*-value	*Lt*	*t*-value
Marijuana						
Between	0	1	2.81	4.17	4.73	3.85
Within	0	1	1.75	10.70	2.13	9.47
Cigarettes						
Between	0	1	1.88	5.85	3.81	5.50
Within	0	1	1.85	10.40	2.92	8.88
Alcohol						
Between	0	1	1.51	5.49	3.10	5.27
Within	0	1	1.77	10.48	2.69	9.71

Table 4
Between-level Correlations Among the Substances

	Marijuana Use		Cigarette Use		Alcohol Use	
	Intercept	Slope	Intercept	Slope	Intercept	Slope
Marijuana Intercept	1.00					
Marijuana Slope	-.25	1.00				
Cigarette Intercept	.70*	.07	1.00			
Cigarette Slope	-.15	.62*	-.21	1.00		
Alcohol Intercept	.65*	-.13	.57*	-.18	1.00	
Alcohol Slope	-.64*	.68*	-.16	.55*	-.54*	1.00

* denotes correlations significant at $p < .05$ or greater.

Table 5
Within-level Correlations Among the Substances

	Marijuana Use		Cigarette Use		Alcohol Use	
	Intercept	Slope	Intercept	Slope	Intercept	Slope
Marijuana Intercept	1.00					
Marijuana Slope	-.20*	1.00				
Cigarette Intercept	.52*	-.10*	1.00			
Cigarette Slope	-.15*	.46*	-.23*	1.00		
Alcohol Intercept	.42*	.04	.35*	.04	1.00	
Alcohol Slope	-.08*	.17*	-.20*	.15*	-.33*	1.00

* denotes correlations significant at $p < .05$ or greater.

individuals who use greater amounts of one substance are more likely to use more of another substance.

Considering the within-level predictors, age had significant linear and quadratic effects on the within-level intercept factors for cigarettes and marijuana. This pattern of effects suggests that within families, the youngest and oldest family members had the lowest initial levels of these substances. Follow-up analyses revealed that peak initial levels of use occurred between the approximate ages of 25-30, corresponding with the age of older siblings and younger parents. No linear or quadratic effect of age was evident for alcohol use.

Significant linear and quadratic effects for age on the growth rate factors were obtained for all three substances. This pattern of effects suggests that the only family members experiencing substantial positive growth in substance use were those family members under the age of 25.

For gender, a significant within-level effect existed for the intercept factor of alcohol use, suggesting higher levels of initial alcohol use for males compared to females.

Given that substantial heterogeneity exists in both intercepts and growth rates across families, it is of interest to investigate how much of this heterogeneity can be further accounted for by certain family-level variables. The family-level variables may be seen as influencing the family-level part of the individual family members' substance use scores indirectly through the family-level factor components. The hierarchical data model with both individual-level and family-level predictors is shown in Figure 3 (next page). The analysis produced a model fit of $\chi^2(185, N = 1204) = 484.13$, $p < .001$, NNFI = .979, CFI = .985.

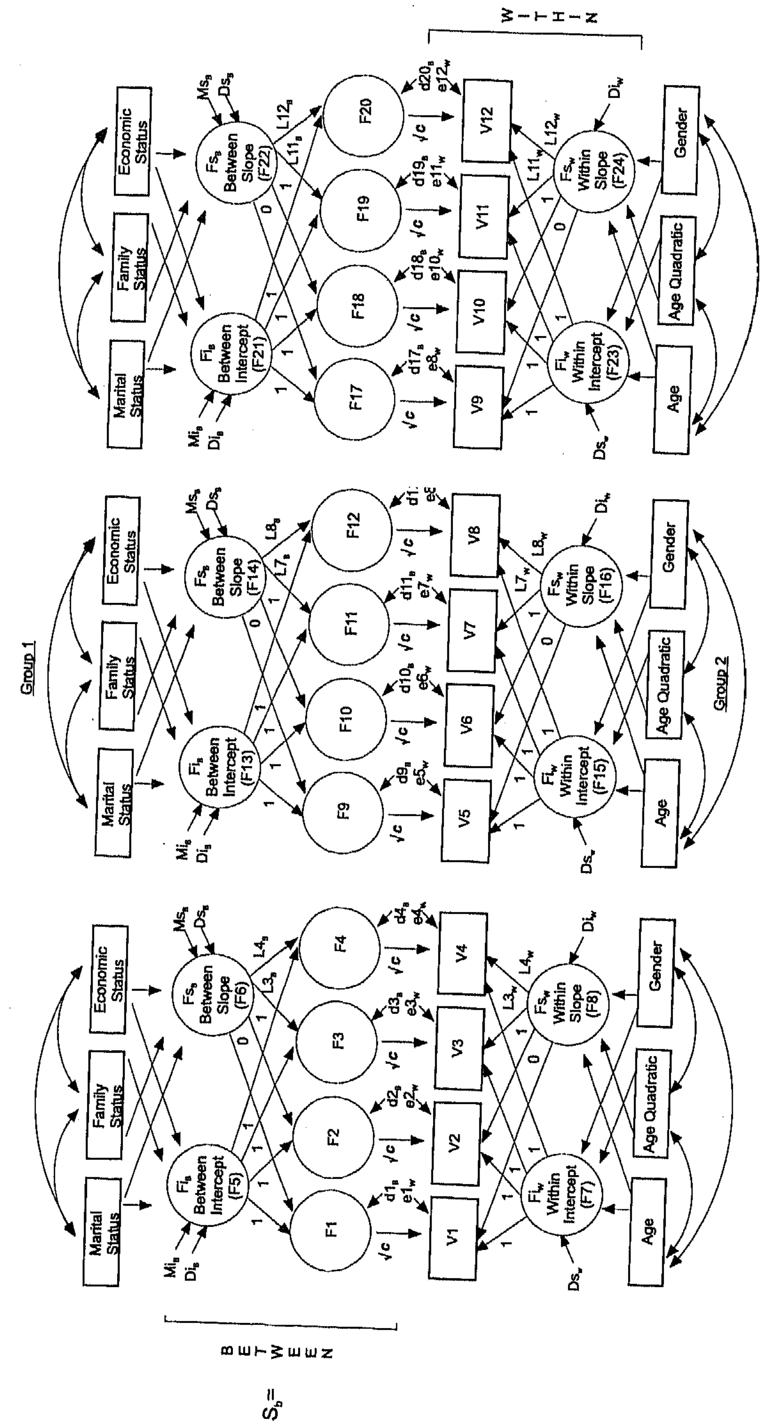
The family-level predictor, marital status, was a significant predictor of the between-level variation in the initial status of both marijuana, -.32, $t = -7.08$, $p < .01$, and alcohol use, -.22, $t = -4.26$, $p < .01$. Family status was a significant predictor of the between-level variation in the initial status of marijuana use, -.32, $t = -4.08$, $p < .01$, cigarettes, -.32, $t = -2.07$, $p < .05$, and alcohol use, -.22, $t = -2.27$, $p < .05$. Socioeconomic status was a significant predictor of the between- level variance in initial status for cigarette use, -.35, $t = -7.23$, $p < .01$, and in the growth rate of marijuana use, -.25, $t = -1.97$, $p < .05$. Combined, these effects accounted for approximately 17.8%, 18.0, and 5.9% of the variation in the family level intercepts for marijuana, cigarettes, and alcohol use, respectively. Socioeconomic status accounted for approximately 4% of the variation in the family level growth rate for marijuana.

In general, single and step-parent families had higher family-levels of substance use, and those families characterized as more economically disadvantaged had higher levels of cigarette use and increased in their use of marijuana at a higher rate. These direct effects on the corresponding between-level substance use factor also suggest that the predictors influenced all of the between-level substance use variables indirectly through the family-level factor.

Discussion

The hierarchical data analysis presented here utilized a latent variable model that not only takes the hierarchical nature of the data into account, but is able to uncover family-level relationships that would be ignored or go undetected due to insufficient power by more conventional analyses. In preliminary analyses sufficient homogeneity in substance use within families was found to warrant a hierarchical analysis, with the proportion of between variation, or intraclass correlation, for the four repeated measures, ranging from .26 to .30 for alcohol, .26 to .28 for cigarettes, and .27 to .33 for marijuana. This made it reasonable to proceed with the multilevel analysis where heterogeneity in substance use among families was examined.

From a substantive point of view, results revealed a significant upward trend in the development of alcohol, cigarette, and marijuana use among families. This finding is consistent with other developmental studies (e.g., Duncan & Duncan, 1994; Duncan, Tildesley, et al., 1995) which have assessed the developmental nature of substance use among adolescents.



WITHIN
Group 1
Group 2
Marital Status
Family Status
Economic Status
Between Intercept (F5)
Between Slope (F6)
Within Intercept (F7)
Within Slope (F8)
Between Intercept (F13)
Between Slope (F14)
Within Intercept (F15)
Within Slope (F16)
Between Intercept (F21)
Between Slope (F22)
Within Intercept (F23)
Within Slope (F24)
Age
Age Quadratic
Gender
F1
F2
F3
F4
F9
F10
F11
F12
F17
F18
F19
F20
V1
V2
V3
V4
V5
V6
V7
V8
V9
V10
V11
V12
BETWEEN
S_b=

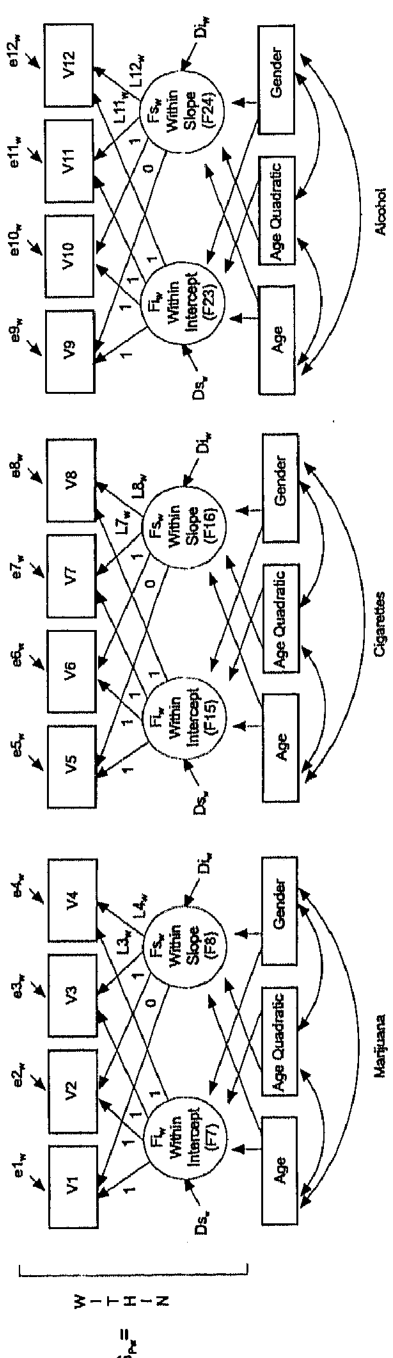


Figure 3
(Divided across pages) Representation of the Hypothesized Multi-level Associative Two-factor LGM Model Using Family-level Covariates
Note. Correlations among between-level growth parameters and within-level growth parameters, presented in Figure 1, are not shown for the sake of clarity.

Initial levels of use were strongly interrelated, indicating that, in general, use of one substance at a higher level is likely to be reflective of use of other substances at elevated levels. The shapes of the family-level developmental curves were also strongly related, indicating that increases in use of the different substances progress at similar rates over the four-year period. This finding supports other literature (e.g., Duncan, Tildesley, et al., 1995; Hansen et al., 1987) which has documented similarities in the developmental trajectories of the three substance use behaviors. The results of the present study add to the existing literature in that they document the developmental association between alcohol, cigarettes, and marijuana, in families, rather than simply among adolescents.

As expected, the within-level effect of age on the developmental trajectories of the three substances revealed that younger family members had higher initial levels of use and developed in their use of the substances at a faster rate than older family members. The within-level effect of gender on the alcohol intercept indicated that males had higher levels of alcohol usage than females, which is consistent with gender effects in other studies (e.g., Donovan & Jessor, 1978; Duncan et al., 1994; Sutker, McCleary, & Allain, 1986).

The between-family variation in substance use scores could be accounted for by family-level covariates. Two contextual variables, or family-level variables, were used to explain the heterogeneity in substance use among families. Consistent with other studies investigating substance use and/or deviant behavior, the specific family-level variables included marital status, family status, and socio-economic status (e.g., Patterson, 1982; Rowe & Gully, 1992; van den Oord, Boomsma, & Verhulst, 1994). Research on the impact of contextual variables such as these on substance use is somewhat equivocal. Although some evidence supports the influence of marital status, education, economic disadvantage, and parent-sibling relationships on substance use and other problem behavior (Duncan et al., 1994; Needle et al., 1986; Rowe & Gully, 1992; van den Oord et al., 1994), this research has primarily been conducted utilizing a target individual as the dependent variable, rather than the family unit as a whole. Results from the present study suggest that, in general, single and step parent families had higher family-levels of substance use, whereas families characterized as less educated and more economically disadvantaged had higher levels of cigarette use and developed in their use of marijuana at a greater rate.

Single parents, particularly mothers, have been shown to have offspring with higher levels of substance use (Byram & Fly, 1984; Hops et al., 1990). We showed this effect in both cross-sectional and longitudinal studies (Hops et al., 1990). Similarly, Burnside, Baer, McLaughlin, & Pokorny (1986) found higher levels of substance use by both adolescents and parents in non-intact families and a significant relation between parent and adolescent alcohol use. Economic disadvantage may also reflect more community-wide influences such as neighborhood disorganization, higher levels of adult crime, and illegal drug trafficking (Hawkins, Catalano, & Miller, 1992).

The problematic behaviors of individuals in the same family may be more alike than that of individuals from different families. With respect to substance use or other problem behavior, there are a number of mechanisms that might account for the commonalities observed among family members. These include social modeling (Akers & Cochran, 1985; Bandura & Walters, 1963) by parents or siblings which may also signify tacit acceptance, the influence and/or encouragement of older siblings (Ary et al., 1993; Brook, Whiteman, Gordon, & Brenden, 1983; Brook, Whiteman, Gordon, Nomura, & Brook, 1986), the availability of illicit substances in the home, increasing the likelihood that initiation or experimental use could occur, and disrupted parenting practices or lax control as a result of substance use by parents (Conger, Reuter, & Conger, 1994; Dishion & Loeber, 1985).

This phenomenon may also partially reflect genetic predispositions to certain behaviors within families with a high degree of consanguinity, or to common environmental and behavioral factors leading to an increased prevalence of specific behaviors within certain families. Clearly, estimates of family level clustering, or other types of clustering (e.g., community level) of specific problem behaviors, such as substance use, can provide additional insights into the etiology and risk factors operating within each social structure.

Using statistical techniques such as LGM provides an opportunity to extend and refine our investigations of the development of adolescent problem behaviors, the context in which they occur, and the antecedents and sequelae of change in such behaviors across the lifespan.

Employing a latent growth curve approach to examine developmental trends in adolescent, parent, and older sibling substance use across a 3-year period, Duncan et al. (1996) investigated similarities in substance use behavior among family members, and the predictive effect of these socializing agents on adolescent use two years later. Results indicated that although both parents and siblings contributed to level of use, only siblings appeared to contribute to the adolescent's subsequent development in substance use. Although the adolescent's developmental trajectory in use was the best predictor of later use, siblings contributed to later use indirectly through their influence on the adolescent's substance use development, while the parent's contribution was more indirect through their influence on the substance use development of the older sibling. These findings suggest that changes in these behaviors are interrelated, and offer support for the interdependent influence of the family.

The basic latent variable growth curve approach advocated here allows for an integrated approach to modeling growth and development that includes both multiple measures and multiple occasions. The approach makes available to a wide audience of researchers the possibility for a variety of analyses of growth and developmental processes. We believe that the potential for integrating typical causal model features found in a majority of SEM applications, and the dynamic features of the latent growth method described here, will make it possible to more precisely understand the influence of the social context on the development of substance use in the family.

The analysis of data that has a hierarchical structure and contains measurements from different levels of the hierarchy requires techniques that are based on assumptions which are in agreement with the data structure. Researchers have struggled for some time with concepts such as hierarchically nested observations, intra-class correlation, the unit of analysis, and random rather than fixed effects. Not only are the more traditional fixed effects analytical methods (e.g., ANOVA) limited in their treatment of the technical difficulties posed by nested designs, but they are also limited in the questions they are able to address. Hierarchical models also represent a useful extension of the traditional variance component models discussed by Winer (1971) and Searle, Casella, and McCulloch (1992), and offer the possibility of making use of within cluster differences in parameter estimates, treating it as a meaningful source of variance rather than as within group error or as a nuisance parameter (Kreft, 1992).

Appropriate techniques are now widely available for univariate response models such as multiple regression through the use of random coefficient or multilevel regression models (e.g., Raudenbush & Bryk, 1988). Extensions of these techniques for multivariate response models (e.g., factor analysis on multiple levels) are only now emerging (Muthén, 1991), even though the substantive concerns about different structures for the between, within, and total covariance matrices gathered for hierarchically nested data are relatively old (see Cronbach, 1976; Burstein, 1980; Harnqvist, 1978).

The present study demonstrated how covariance structure models can be formulated for hierarchical data and how they can be analyzed with conventional SEM software by means of a simple ad hoc estimator. Instructions for the use of the program utilized in the present study are given in Nelson and Muthén (1991). Several extensions of the basic multilevel covariance structure model are possible (see Schmidt & Wisenbaker, 1986; Muthén, 1989), and these extensions can be readily carried out utilizing conventional SEM software and the estimator provided by Muthén (1989).

The flexibility of the multilevel covariance structure model makes it an attractive analytic tool for a variety of SEM analyses, including the use of latent growth models, to investigate growth and development among variables of interest with multilevel data. Within the specific area of adolescent substance use and problem behaviors, hierarchical latent variable models allow for potentially greater insight into the developmental nature, antecedents, and sequelae of a variety of adolescent problem behaviors.

References

Aitkin, M. & Longford, N. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, Series A, 149, 1*, 1-43.

Akers, R. L. & Cochran, J. E. (1985). Adolescent marijuana use: A test of three theories of deviant behavior. *Deviant Behavior, 6*, 323-346.

Ary, D. V., Tildesley, E., Hops, H., & Andrews, J. (1993). The influence of parent, sibling, and peer modeling and attitudes on adolescent use of alcohol. *International Journal of the Addictions, 28*, 853-880.

Bachman, J. G., Johnston, L. D., & O'Malley, P. M. (1981). Smoking, drinking, and drug use among American high school students: Correlates and trends, 1975-1979. *American Journal of Public Health, 71*, 59-69.

Bandura, A. & Walters, R. H. (1963). *Social learning and personality development.* New York: Holt, Rinehart and Winston.

Bell, R. Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review, 75*, 81-95.

Bentler, P. M. (1990). *EQS: Structural equations program manual.* Los Angeles: BMDP Statistical Software.

Berndt, T. J. (1982). The features and effects of friendships in early adolescence. *Child Development, 53*, 1447-1460.

Biernat, M. & Wortman, C. (1991). Sharing of home responsibilities between professionally employed women and their husbands. *Journal of Personality and Social Psychology, 60*, 844- 860.

Biglan, A., Metzler, C. W., Wirt, R., Ary, D., Noell, J., Ochs, L., French, C., & Hood, D. (1990). Social and behavioral factors associated with high-risk sexual behavior among adolescents. *Journal of Behavioral Medicine, 13*, 245-261.

Brennan, T. (1979). *Patterns of multiple drug use in a national adolescent sample: A multivariate approach.* Boulder, CO: Behavioral Research Institute.

Bronfrenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 3*, 513-531.

Brook, J. S., Whiteman, M., Gordon, A. S., & Brenden, C. (1983). Older brother's influence on younger sibling's drug use. *Journal of Psychology, 114*, 83-90.

Brook, J. S., Whiteman, M., Gordon, A. S., Nomura, C., & Brook, D. W. (1986). Onset of adolescent drinking: A longitudinal study of intrapersonal and interpersonal antecedents. In B. Stimmel (Ed.), *Alcohol and substance abuse in women and children* (pp. 91-110). Binghamton, NY: Haworth Press.

Bryk, A. S. & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*, 147-158.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Burchinal, M. & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development*, *62*, 23-43.

Burnside, M. A., Baer, P. E., McLaughlin, R. J., & Pokorny, A. D. (1986). Alcohol use by adolescents in disrupted families. *Alcoholism: Clinical and Experimental Research*, *10*, 274- 278.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. *Review of Research in Education*, *8*, 158-233.

Byram, O. W. & Fly, J. W. (1984). Family structure, race, and adolescents' alcohol use: A research note. *American Journal of Drug and Alcohol Abuse*, *10*, 467-478.

Conger, R. E., Reuter, M. A., & Conger, K. J. (1994). The family context of adolescent vulnerability and resilience to alcohol use and abuse. *Sociological Studies of Children*, *6*, 55-86.

Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis*. Unpublished manuscript, Stanford University, Stanford Evaluation Consortium, School of Education.

Curran, P. J., Harford, T., & Muthén, B. O. (1996). The relation between heavy alcohol use and bar patronage: A latent growth model. *Journal of Studies on Alcohol*, *57*, 410-418.

de Leeuw, J. & Kreft, I. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics*, *11*, 57-85.

Dishion, T. J. & Loeber, R. (1985). Male adolescent marijuana and alcohol use: The role of parents and peers revisited. *American Journal of Drug and Alcohol Abuse*, *11*, 11-25.

Donovan, J. E. & Jessor, R. (1978). Adolescent problem drinking: Psychosocial correlates in a national sample study. *Journal of Studies on Alcohol*, *39*, 1506-1524.

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

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

Duncan, S. C. & Duncan, T. E. (1994). Modeling incomplete longitudinal substance use data using latent variable growth curve methodology. *Multivariate Behavioral Research*, *29*, 313- 338.

Duncan, T. E. & Duncan, S. C. (1995). Modeling the processes of development via latent growth curve methodology. *Structural Equation Modeling*, *2*, 187-213.

Duncan, T. E., Duncan, S. C., & Hops, H. (1994). The effects of family cohesiveness and peer encouragement on the development of adolescent alcohol use: A cohort-sequential approach to the analysis of longitudinal data. *Journal of Studies on Alcohol*, *55*, 588-599.

Duncan, T. E., Duncan, S. C., & Hops, H. (1996). The role of parents and older siblings in predicting adolescent substance use: Modeling development via structural equation latent growth curve methodology. *Journal of Family Psychology*, *10*, 158-172.

Duncan, T. E., Duncan, S. C., Hops, H., & Stoolmiller, M. (1995). Analyzing longitudinal substance use data via generalized estimating equation methodology. *Multivariate Behavioral Research*, *30*, 317-339.

Duncan, T. E. & McAuley, E. (1993). Social support and efficacy cognitions in exercise adherence: A latent growth curve analysis. *Journal of Behavioral Medicine*, *16*, 199-218.

Duncan, T. & Stoolmiller, M. (1993). Modeling social and psychological determinants of exercise behaviors via structural equation systems. *Research Quarterly for Exercise and Sport*, *64*, 1-16.

Duncan, T. E., Tildesley, E. A., Duncan, S. C., & Hops, H. (1995). The consistency of family and peer influence on the development of substance use in adolescence. *Addiction*, *90*, 1647- 1660.

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.

Gfroerer, J. (1987). Correlations between drug use by teenagers and drug use by older family members. *American Journal of Drug and Alcohol Abuse*, *13*, 95-108.

Goldstein, H. I. (1986). Multilevel mixed linear model analysis using iterative general least squares. *Biometrika*, *73*, 43-56.

Hansen, W. B., Graham, J. W., Sobel, J. L., Shelton, D. R., Flay, B. R., & Johnson, C. A. (1987). The consistency of peer and parent influences on tobacco, alcohol, and marijuana use among young adolescents. *Journal of Behavioral Medicine*, *10*, 559-579.

Harnqvist, K. (1978). Primary mental abilities of collective and individual levels. *Journal of Educational Psychology*, *70*, 706-716.

Hawkins, J. D., Catalano, R., & Miller, J. (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.

Hawkins, J. D., Lishner, D. M., Catalano, R. F., & Howard, M. D. (1986). Childhood predictors and the prevention of adolescent substance abuse: Toward an empirically grounded theory. *Journal of Children in Contemporary Society*, *18*, 11-48.

Hollingshead, A. B. (1975). *Four-factor index of social status*. New Haven, CT: Yale University Working Paper.

Hops, H., Duncan, T. E., Duncan, S. C., & Stoolmiller, M. (1996). Parent substance use as a predictor of adolescent use: A six-year lagged analysis. *Annals of Behavioral Medicine*, *18*, 157- 164.

Hops, H., Tildesley, E., Lichtenstein, E., Ary, D., & Sherman, L. (1990). Parent-adolscent problem-solving interactions and drug use. *American Journal of Drug and Alcohol Abuse*, *16*, 239-258.

Jessor, R. (1987a). Problem-behavior theory, psychosocial development, and adolescent problem drinking. *British Journal of Addiction*, *82*, 331-342.

Jessor, R. (1987b). Risky driving and adolescent problem behavior: An extension of problem-behavior theory. *Alcohol, Drugs, and Driving*, *3(3-4)*, 1-11.

Jessor, R. & Jessor, S. L. (1977). *Problem behavior and psychosocial development*. New York: Academic Press.

Johnston, L. D., O'Malley, P. M. & Bachman, J. G. (1993). *National survey results on drug use from the Monitoring the Future Study, 1975-1992*. Rockville, MD: National Institute on Drug Abuse.

Keesling, J. W. & Wiley, D. E. (1974, March). *Regression models of hierarchical data*. Paper presented at the Annual Meeting of the Psychometric Society, Palo Alto, CA.

Koch, G. G. (1983). Intraclass correlation coefficient. *Encyclopedia of Statistical Sciences*, *4*, 212-217.

Kreft, I. G. (1992). Multilevel models for hierarchically nested data: Potential applications in substance abuse prevention research. In L. Collins & L. Seitz (Eds.), *Technical review panel on advances in data analysis for prevention intervention research*. NIDA Research Monograph, 108.

Loeber, R. & Dishion, T. (1983). Early predictors of male delinquency: A review. *Psychological Bulletin*, *94*, 68-99.

Longford, N. T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. *Biometrika*, *74*, 817-827.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In R. B. Cattel & J. Nesselroade (Eds.), *Handbook of multivariate experimental psychology*, 2nd ed. (pp. 561-614). New York: Plenum Press.

McArdle, J. J. & Anderson, E. R. (1989). Latent growth models for research on aging. In L. E. Biren & K. W. Schaie (Eds.), *The handbook of the psychology of aging*, 3rd ed. (pp. 21-44). San Diego: Academic Press.

McArdle, J. J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110-133.

McArdle, J. J. & Hamagami, F. (1991). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. In L. M. Collins & J. C. Horn (Eds.), *Best methods for the analysis of change* (pp. 276-304). Washington, DC: American Psychological Association.

Melby, J. N., Conger, R. D., Conger, K. J., & Lorenz, F. O. (1993). Effects of parental behavior on tobacco use by young male adolescents. *Journal of Marriage and the Family*, *55*, 439-454.

Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107-122.

Minuchin, S. (1974). *Families and family therapy*. Cambridge, MA: Harvard University Press.

Montemayor, R. & Brownlee, J. R. (1987). Fathers, mothers, and adolescents: Gender-based differences in parental roles during adolescence. *Journal of Youth and Adolescence*, *16*, 281-291.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. Presidential address to the Psychometric Society. *Psychometrika*, *54*, 557-585.

Muthén, B. (1990, June). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society Meeting in Princeton, NJ.

Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, *28*, 338-354.

Muthén, B. (1992, June). *Mean and covariance structure analysis of hierarchical data*. Paper presented at the Psychometric Society meeting in Princeton, NC.

Muthén, B. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376-398.

Muthén, B. (1997). Latent variable modeling of longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological Methodology* (pp. 453-480). Boston: Blackwell.

Muthén, B. & Satorra, A. (1989). Multilevel aspects of varying parameters in structural models. In R. D. Bock (Ed.), *Multilevel analysis of educational data* (pp. 87-99). San Diego, CA: Academic Press.

National Institute on Drug Abuse (1989). *Overview of the 1988 National Household Survey on Drug Abuse*. Rockville, MD: Author.

Needle, R., McCubbin, H., Wilson, M., Reineck, R., Lazar, A., & Mederer, H. (1986). Interpersonal influences in adolescent drug use — The role of older siblings, parents and peers. *International Journal of the Addictions*, *21*, 739-766.

Nelson, G. & Muthén, B. (1991). *Analysis preparation steps for multilevel analysis using LISCOMP* (Technical Report). Los Angeles: University of California.

Newcomb, M. D. & Bentler, P. M. (1988). Impact of adolescent drug use and social support on problems of young adults: A longitudinal study. *Journal of Abnormal Psychology*, *97(1)*, 64-75.

Oregon Employment Division Research and Statistics (1989). *Is Oregon's future at risk? A profile of Oregon's youth*. Salem, OR: Author.

Osgood, D. W., Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (1988). The generality of deviance in late adolescence and early adulthood. *American Sociology Review*, *53*, 81-93.

Patterson, G. R. (1982). *A social learning approach: III. Coercive family process*. Eugene, OR: Castalia Publishing Company.

Patterson, G. R. (1984). Siblings: Fellow travelers in a coercive system. In R. J. Blanchard & D. C. Blanchard (Eds.), *Advances in the study of aggression* (Vol. 1, pp. 173-215). New York: Academic.

Patterson, G. R. (1986). The contribution of siblings to training for fighting: A microsocial analysis. In D. Olweus, J. Block, & M. Radke-Yarow (Eds.), *Development of antisocial and prosocial behaviors* (pp. 235-260). Orlando, FL: Academic Press.

Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics, 14*, pp. 1-17.

Raudenbush, S. & Bryk, A. (1988). Methodological advances in studying effects of schools and classrooms on student learning. In E. Z. Roth (Ed.), *Review of research in education* (pp. 423- 475). Washington, DC: American Educational Research Association.

Robbins, C. A. & Martin, S. S. (1993). Gender, styles of deviance, and drinking problems. *Journal of Health and Social Behavior, 34*, 302-321.

Rowe, D. C. & Gulley, B. L. (1992). Sibling effects of substance use and delinquency. *Criminology, 30*, 217-233.

Schmidt, W. & Wisenbaker, J. (1986). *Hierarchical data analysis: an approach based on structural equations* (Tech. Rep. No. 4.). East Lansing, MI: Michigan State University, Department of Counseling Educational Psychology and Special Education.

Searle, S. R., Casella, G., & McCulloch., C. E. (1992). *Variance components.* New York: Wiley & Sons.

Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: Growth and change from childhood to early adolescence. *Multivariate Behavioral Research, 29*, 263-288.

Stoolmiller, M., Duncan, T., Bank, L., & Patterson, G. (1993). Some problems and solutions in the study of change: Significant patterns in client resistance. *Journal of Clinical and Consulting Psychology, 61*, 920-928.

Sutker, P. B., McCleary, G. E., & Allain, A. N. (1986). Adolescent alcohol abuse. In R. A. Feldman & A. R. Stiffman (Eds.), *Advances in adolescent mental health*, Vol. 1 (pp. 195-223). Greenwich, CT: JAI Press.

Tildesley, E., Hops, H., Ary, D., & Andrews, J. A. (1995). A multitrait-multimethod model for adolescent deviance, drug use, academic, and sexual behaviors. *Journal of Psychopathology and Behavioral Assessment, 17*, 185-215.

Tisak, J. & Meredith, W. (1990). Descriptive and associative developmental models. In A. von Eye (Ed.), *New statistical methods in developmental research.* New York: Academic Press.

Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika, 23*, 19-23.

van den Oord, J. C. G., Boomsma, D. I., & Verhulst, F. C. (1994). A study of problem behaviors in 10- to 15-year-old biologically related and unrelated international adoptees. *Behavior Genetics, 24*, 193-205.

Willet, J. B., Ayoub, C. C., & Robinson, D. (1991). Using growth modeling to examine systematic differences in growth: An example of change in the function of families at risk of maladaptive parenting, child abuse, or neglect. *Journal of Consulting and Clinical Psychology, 59*, 38-47.

Willet, J. B. & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363-381.

Winer, B. J. (1971). *Statistical principles in experimental design.* New York: McGraw-Hill.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design.* New York: McGraw-Hill.

Zabin, L. (1984). The association between smoking and sexual behavior among teens in U.S. contraceptive clinics. *American Journal of Public Health*, *74*, 261-263.

Accepted May, 1997

Appendix A

Specification for the Latent Variable Modeling of the Longitudinal and Multilevel Substance Use Variables

The following program specifications and input data correspond to those necessary to test the hypothesized model (Figure 1) utilizing the EQS (version 5.1) structural equations program. Although the SOURCEBW program produces data in a rectangular format, EQS requires data formatted in lower triangular matrices. A program that converts the output is available upon request from the authors.

Note that some lines wrap to a second line. For these, the second line has been aligned on the right margin.

```
/TITLE
 MULTISAMPLE LONGITUDINAL ANALYSIS GROUP 1
/SPECIFICATIONS
 CAS=435; VAR=15; ME=ML; FI=4; MA=COV; ANAL=MOMENT;
GROUPS=2;
/LABELS
 V1=POT_T1; V2=POT_T2; V3=POT_T3; V4=POT_T4;
 V5=CIG_T1; V6=CIG_T2; V7=CIG_T3; V8=CIG_T4;
 V9=ALC_T1; V10=ALC_T2; V11=ALC_T3; V12=ALC_T4;
 V13=SEX; V14=AGE; V15=AGESQ;
 F5=BPOT_INT; F6=BPOT_SLP;
 F7=WPOT_INT; F8=WPOT_SLP;
 F13=BCIG_INT; F14=BCIG_SLP;
 F15=WCIG_INT; F16=WCIG_SLP;
 F21=BALC_INT; F22=BALC_SLP;
 F23=WALC_INT; F24=WALC_SLP;
/EQUATIONS
 V1=1.6635F1+F7+0F8+E1;
 V2=1.6635F2+F7+1F8+E2;
 V3=1.6635F3+F7+2.105*F8+E3;
 V4=1.6635F4+F7+2.961*F8+E4;
```

```
F1=F5+0F6+D1;
F2=F5+1F6+D2;
F3=F5+3.821*F6+D3;
F4=F5+7.139*F6+D4;
F5=1.577*V999+D5;
F6=.034*V999+D6;
F7=0V999+*V13+*V14+*V15+D7;
F8=0V999+*V13+*V14+*V15+D8;
V5=1.6635F9+F15+0F16+E5;
V6=1.6635F10+F15+1F16+E6;
V7=1.6635F11+F15+2.006*F16+E7;
V8=1.6635F12+F15+3.621*F16+E8;
F9=F13+0F14+D9;
F10=F13+1F14+D10;
F11=F13+1.725*F14+D11;
F12=F13+3.830*F14+D12;
F13=2.015*V999+D13;
F14=.118*V999+D14;
F15=0V999+*V13+*V14+*V15+D15;
F16=0V999+*V13+*V14+*V15+D16;
V9=1.6635F17+F23+0F24+E9;
V10=1.6635F18+F23+1F24+E10;
V11=1.6635F19+F23+1.651*F24+E11;
V12=1.6635F20+F23+2.588*F24+E12;
F17=F21+0F22+D17;
F18=F21+1F22+D18;
F19=F21+1.725*F22+D19;
F20=F21+3.830*F22+D20;
F21=2.015*V999+D21;
F22=.118*V999+D22;
F24=0V999+*V13+*V14+*V15+D24;
F23=0V999+*V13+*V14+*V15+D23;
V13=0*V999+E13;
V14=0*V999+E14;
V15=0*V999+E15;
 /VARIANCES
E13 TO E15=*;
E1=.230*; E2=.188*; E3=.155*; E4=.386*;
E5=.185*; E6=.230*; E7=.235*; E8=.392*;
E9=.396*; E10=.295*; E11=.312*; E12=.446*;
D1 TO D4=*;
```

```
 D5 TO D6=*;
 D7=.493*; D8=.024*;
 D9 TO D12=*;
 D13 TO D14=*;
 D15=.969*; D16=.026*;
 D17 TO D20=*;
 D21 TO D22=*;
 D23=.611*; D24=.044*;
/COVARIANCES
 E13 TO E15=*;
 D5 TO D6=*;
 D13 TO D14=*;
 D21 TO D22=*;
 D5,D13=*;
 D5,D14=*;
 D5,D21=*;
 D5,D22=*;
 D6,D13=*;
 D6,D14=*;
 D6,D21=*;
 D6,D22=*;
 D13,D21=*;
 D13,D22=*;
 D14,D21=*;
 D14,D22=*;
 D7 TO D8=*;
 D15 TO D16=*;
 D23 TO D24=*;
 D7,D15=*;
 D7,D16=*;
 D7,D23=*;
 D7,D24=*;
 D8,D15=*;
 D8,D16=*;
 D8,D23=*;
 D8,D24=*;
 D15,D23=*;
 D15,D24=*;
 D16,D23=*;
 D16,D24=*;
 E1,E5=*; E2,E6=*; E3,E7=*; E4,E8=*;
```

```
 E1,E9=*; E2,E10=*; E3,E11=*; E4,E12=*;
 E5,E9=*; E6,E10=*; E7,E11=*; E8,E12=*;
/CONSTRAINTS
/MEANS
 0.263341D+01 0.273427D+01 0.288763D+01 0.303271D+01
 0.327726D+01 0.343753D+01 0.352457D+01 0.378156D+01
 0.332248D+01 0.346387D+01 0.356572D+01 0.377803D+01
 -.110531D-03 0.967150D-04 -.118821D-02
/MATRIX
 0.123244D+01
 0.102526D+01 0.130851D+01
 0.102334D+01 0.122456D+01 0.155962D+01
 0.833812D+00 0.100101D+01 0.128312D+01 0.156030D+01
 0.924670D+00 0.905813D+00 0.937615D+00 0.842007D+00
 0.869099D+00 0.970173D+00 0.102154D+01 0.966174D+00
 0.808962D+00 0.979154D+00 0.112769D+01 0.105422D+01
 0.738267D+00 0.874639D+00 0.109239D+01 0.122831D+01
 0.754062D+00 0.729507D+00 0.788402D+00 0.650722D+00
 0.558696D+00 0.676777D+00 0.752982D+00 0.690159D+00
 0.491373D+00 0.621703D+00 0.764011D+00 0.672423D+00
 0.372754D+00 0.532511D+00 0.672357D+00 0.786896D+00
 0.715222D-01 0.535610D-01 0.285135D-01 0.667851D-03
 -.394120D-01 -.105655D+00 -.197636D+00 -.284920D+00
 -.817747D-01 -.150956D+00 -.241460D+00 -.322713D+00
 0.212129D+01
 0.190590D+01 0.217070D+01
 0.181592D+01 0.194424D+01 0.224808D+01
 0.165197D+01 0.184806D+01 0.202640D+01 0.253606D+01
 0.845977D+00 0.838329D+00 0.820287D+00 0.782263D+00
 0.717859D+00 0.813592D+00 0.827683D+00 0.800729D+00
 0.581530D+00 0.677362D+00 0.707852D+00 0.754652D+00
 0.575851D+00 0.694680D+00 0.794413D+00 0.999457D+00
 0.116156D+00 0.109174D+00 0.630726D-01 0.590290D-02
 0.350832D-01 -.801662D-01 -.157919D+00 -.276479D+00
 -.464450D-02 -.117511D+00 -.192967D+00 -.305495D+00
 0.176458D+01
 0.134126D+01 0.170268D+01
 0.127132D+01 0.142166D+01 0.174124D+01
 0.105810D+01 0.124478D+01 0.134880D+01 0.180661D+01
 -.461751D-02 0.308984D-01 0.112603D-01 0.649749D-03
 -.125610D+00 -.186117D+00 -.299269D+00 -.433523D+00
```

```
 -.148858D+00 -.202823D+00 -.315204D+00 -.442192D+00
 0.800215D+00
 0.190493D-01 0.679126D+00
 0.502300D-02 0.708155D+00 0.759360D+00
/CONSTRAINTS
/PRINT
EFFECTS=YES;
/END

/TITLE
 MULTISAMPLE LONGITUDINAL ANALYSIS GROUP 2
/SPECIFICATIONS
 CAS=769; VAR=15; ME=ML; FI=4; MA=COV; ANAL=MOMENT;
/LABELS
 V1=POT_T1; V2=POT_T2; V3=POT_T3; V4=POT_T4;
 V5=CIG_T1; V6=CIG_T2; V7=CIG_T3; V8=CIG_T4;
 V9=ALC_T1; V10=ALC_T2; V11=ALC_T3; V12=ALC_T4;
 V13=SEX; V14=AGE; V15=AGESQ;
 F5=BPOT_INT; F6=BPOT_SLP;
 F7=WPOT_INT; F8=WPOT_SLP;
 F13=BCIG_INT; F14=BCIG_SLP;
 F15=WCIG_INT; F16=WCIG_SLP;
 F21=BALC_INT; F22=BALC_SLP;
 F23=WALC_INT; F24=WALC_SLP;
/EQUATIONS
 V1=F7+0F8+E1;
 V2=F7+1F8+E2;
 V3=F7+*F8+E3;
 V4=F7+*F8+E4;
 F7=0V999+*V13+*V14+*V15+D7;
 F8=0V999+*V13+*V14+*V15+D8;
 V5=F15+0F16+E5;
 V6=F15+1F16+E6;
 V7=F15+*F16+E7;
 V8=F15+*F16+E8;
 F15=0V999+*V13+*V14+*V15+D15;
 F16=0V999+*V13+*V14+*V15+D16;
 V9=F23+0F24+E9;
 V10=F23+1F24+E10;
 V11=F23+*F24+E11;
 V12=F23+*F24+E12;
```

```
 F23=0V999+*V13+*V14+*V15+D23;
 F24=0V999+*V13+*V14+*V15+D24;
 V13=0V999+E13;
 V14=0V999+E14;
 V15=0V999+E15;
/VARIANCES
 E1 TO E4=*;
 E5 TO E8=*;
 E9 TO E12=*;
 E13 TO E15=*;
 D7 TO D8=*;
 D15 TO D16=*;
 D23 TO D24=*;
/COVARIANCES
 E13 TO E15=*;
 D7 TO D8=*;
 D15 TO D16=*;
 D23 TO D24=*;
 D7,D15=*;
 D7,D16=*;
 D7,D23=*;
 D7,D24=*;
 D8,D15=*;
 D8,D16=*;
 D8,D23=*;
 D8,D24=*;
 D15,D23=*;
 D15,D24=*;
 D16,D23=*;
 D16,D24=*;
 E1,E5=*; E2,E6=*; E3,E7=*; E4,E8=*;
 E1,E9=*; E2,E10=*; E3,E11=*; E4,E12=*;
 E5,E9=*; E6,E10=*; E7,E11=*; E8,E12=*;
/MEANS
 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
/MATRIX
 0.597941D+00
 0.451452D+00 0.575098D+00
 0.435999D+00 0.489250D+00 0.628262D+00
 0.376225D+00 0.435197D+00 0.519983D+00 0.671196D+00
 0.404010D+00 0.326463D+00 0.308929D+00 0.246034D+00
```

```
0.348721D+00 0.346489D+00 0.354725D+00 0.301474D+00
0.319289D+00 0.345904D+00 0.381469D+00 0.332596D+00
0.246771D+00 0.299913D+00 0.352146D+00 0.394170D+00
0.249277D+00 0.262131D+00 0.259662D+00 0.260101D+00
0.153449D+00 0.257163D+00 0.254346D+00 0.284574D+00
0.157300D+00 0.246547D+00 0.285089D+00 0.311073D+00
0.124299D+00 0.232078D+00 0.286982D+00 0.373836D+00
0.424005D-01 0.570351D-02 -.238752D-02 -.285618D-01
0.657789D-01 -.643442D-01 -.101414D+00 -.176622D+00
0.375834D-01 -.801972D-01 -.115297D+00 -.184258D+00
0.108229D+01
0.927482D+00 0.108791D+01
0.864976D+00 0.947594D+00 0.108498D+01
0.741071D+00 0.853164D+00 0.931686D+00 0.118394D+01
0.265420D+00 0.291420D+00 0.304452D+00 0.306608D+00
0.114566D+00 0.226331D+00 0.247684D+00 0.305387D+00
0.890017D-01 0.193595D+00 0.246178D+00 0.311714D+00
0.405093D-01 0.168359D+00 0.249378D+00 0.403578D+00
0.127776D-01 -.240520D-01 -.210455D-01 -.742783D-01
0.172401D+00 0.349902D-01 -.353338D-01 -.215484D+00
0.152692D+00 0.237020D-01 -.413264D-01 -.211327D+00
0.856152D+00
0.647885D+00 0.977026D+00
0.659252D+00 0.817337D+00 0.108340D+01
0.608559D+00 0.790688D+00 0.951830D+00 0.129971D+01
-.132539D+00 -.160860D+00 -.176488D+00 -.209352D+00
-.273909D+00 -.461551D+00 -.604660D+00 -.791165D+00
-.271842D+00 -.441232D+00 -.581611D+00 -.755966D+00
0.111472D+01
0.213724D+00 0.118130D+01
0.182660D+00 0.114727D+01 0.113450D+01
/CONSTRAINTS
(1,V3,F8)=(2,V3,F8);
(1,V4,F8)=(2,V4,F8);
(1,E1,E1)=(2,E1,E1);
(1,E2,E2)=(2,E2,E2);
(1,E3,E3)=(2,E3,E3);
(1,E4,E4)=(2,E4,E4);
(1,D7,D7)=(2,D7,D7);
(1,D8,D8)=(2,D8,D8);
(1,D7,D8)=(2,D7,D8);
```

```
(1,V7,F16)=(2,V7,F16);
(1,V8,F16)=(2,V8,F16);
(1,E5,E5)=(2,E5,E5);
(1,E6,E6)=(2,E6,E6);
(1,E7,E7)=(2,E7,E7);
(1,E8,E8)=(2,E8,E8);
(1,D15,D15)=(2,D15,D15);
(1,D16,D16)=(2,D16,D16);
(1,D15,D16)=(2,D15,D16);
(1,V11,F24)=(2,V11,F24);
(1,V12,F24)=(2,V12,F24);
(1,E9,E9)=(2,E9,E9);
(1,E10,E10)=(2,E10,E10);
(1,E11,E11)=(2,E11,E11);
(1,E12,E12)=(2,E12,E12);
(1,D23,D23)=(2,D23,D23);
(1,D24,D24)=(2,D24,D24);
(1,D23,D24)=(2,D23,D24);
(1,D7,D16)=(2,D7,D16);
(1,D8,D16)=(2,D8,D16);
(1,D7,D15)=(2,D7,D15);
(1,D8,D15)=(2,D8,D15);
(1,D7,D23)=(2,D7,D23);
(1,D8,D23)=(2,D8,D23);
(1,D7,D24)=(2,D7,D24);
(1,D8,D24)=(2,D8,D24);
(1,D15,D23)=(2,D15,D23);
(1,D16,D23)=(2,D16,D23);
(1,D15,D24)=(2,D15,D24);
(1,D16,D24)=(2,D16,D24);
(1,E1,E5)=(2,E1,E5);
(1,E1,E9)=(2,E1,E9);
(1,E2,E6)=(2,E2,E6);
(1,E2,E10)=(2,E2,E10);
(1,E3,E7)=(2,E3,E7);
(1,E3,E11)=(2,E3,E11);
(1,E4,E8)=(2,E4,E8);
(1,E4,E12)=(2,E4,E12);
(1,E5,E9)=(2,E5,E9);
(1,E6,E10)=(2,E6,E10);
(1,E7,E11)=(2,E7,E11);
```

```
 (1,E8,E12)=(2,E8,E12);
 (1,E13,E13)=(2,E13,E13);
 (1,E14,E14)=(2,E14,E14);
 (1,E13,E14)=(2,E13,E14);
 (1,E15,E15)=(2,E15,E15);
 (1,E15,E13)=(2,E15,E13);
 (1,E15,E14)=(2,E15,E14);
 (1,F7,V13)=(2,F7,V13);
 (1,F7,V14)=(2,F7,V14);
 (1,F8,V13)=(2,F8,V13);
 (1,F8,V14)=(2,F8,V14);
 (1,F7,V15)=(2,F7,V15);
 (1,F8,V15)=(2,F8,V15);
 (1,F15,V13)=(2,F15,V13);
 (1,F15,V14)=(2,F15,V14);
 (1,F16,V13)=(2,F16,V13);
 (1,F16,V14)=(2,F16,V14);
 (1,F15,V15)=(2,F15,V15);
 (1,F16,V15)=(2,F16,V15);
 (1,F23,V13)=(2,F23,V13);
 (1,F23,V14)=(2,F23,V14);
 (1,F24,V13)=(2,F24,V13);
 (1,F24,V14)=(2,F24,V14);
 (1,F23,V15)=(2,F23,V15);
 (1,F24,V15)=(2,F24,V15);
/PRINT
EFFECTS=YES;
/TECH
ITER=100;
/LMTEST
/END
```