Multivariate Behavioral Research, 43:289–326, 2008

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273170802034893

Psychology Press
Taylor & Francis Group

Using Multilevel Mixtures to Evaluate Intervention Effects in Group Randomized Trials

M. Lee Van Horn
Department of Psychology
University of South Carolina

Abigail A. Fagan
Department of Criminology
University of South Carolina

Thomas Jaki
Department of Mathematics and Statistics
Lancaster University

Eric C. Brown, J. David Hawkins, Michael W. Arthur,
Robert D. Abbott, and Richard F. Catalano
Social Development Research Group
University of Washington

There is evidence to suggest that the effects of behavioral interventions may be limited to specific types of individuals, but methods for evaluating such outcomes have not been fully developed. This study proposes the use of finite mixture models to evaluate whether interventions, and, specifically, group randomized trials, impact participants with certain characteristics or levels of problem behaviors. This study uses latent classes defined by clustering of individuals based on the targeted behaviors and illustrates the model by testing whether a preventive intervention aimed at reducing problem behaviors affects experimental users of illicit substances

Correspondence concerning this article should be addressed to M. Lee Van Horn, Department of Psychology, Barnwell College, University of South Carolina, Columbia, SC 29206. E-mail: vanhorn@sc.edu

differently than problematic substance users or those individuals engaged in more serious problem behaviors. An illustrative example is used to demonstrate the identification of latent classes, specification of random effects in a multilevel mixture model, independent validation of latent classes, and the estimation of power for the proposed models to detect intervention effects. This study proposes specific steps for the estimation of multilevel mixture models and their power and suggests that this model can be applied more broadly to understand the effectiveness of interventions.

Prevention scientists seek to improve health outcomes for youth. Although some prevention programs have been found to successfully prevent adolescent drug use or delinquency (Hawkins, 2004; Loeber & Farrington, 1998; Mihalic, Fagan, Irwin, Ballard, & Elliott, 2004; Sherman et al., 1997; U.S. Department of Health and Human Services, 2001), evaluations of other preventive interventions have found that effects differ across participants' demographic characteristics, such as sex, race/ethnicity, income, and other factors. For example, an evaluation of the Good Behavior Game (Kellam, Ling, Merisca, Brown, & Ialongo, 1998) reported reductions in participants' observed levels of playground aggression compared with those in the control condition, but these effects were significant only for males. The evaluation of the Lion's-Quest Skills for Adolescence (SFA) curriculum indicated reduced alcohol use and binge drinking for Hispanic students in the intervention compared with controls but showed no effects on these outcomes for White students (Eisen, Zellman, Massett, & Murray, 2002). In contrast, evaluations of other prevention programs have not found much evidence for differential effectiveness according to demographic characteristics (Elliott & Mihalic, 2004). For example, the Life Skills Training drug prevention program has shown reductions in substance use for White, middle-class students attending suburban schools as well as minority students attending inner-city schools (Botvin, Mihalic, & Grotpeter, 1998).

Differences in intervention effects also may be related to participants' levels of the behaviors or attitudes the intervention targets for change. Evaluations of some prevention programs indicated larger effects for high-risk participants (i.e., those demonstrating greater levels of targeted risk factors or problem behaviors at pretest; see Allen & Philliber, 2001; Frey et al., 2005; Kellam et al., 1998; Lochman, Curry, Dane, & Ellis, 2001; Stoolmiller, Eddy, & Reid, 2000). In addition to finding overall program effects for the Teen Outreach Program, Allen and Philliber found larger reductions in teenage pregnancy, course failure, and school suspension for intervention students who reported higher baseline levels of teenage pregnancy, school suspension, and school risk factors, respectively. Similarly, the Good Behavior Game demonstrated greater decreases in observed playground aggression for boys who were rated as having the highest levels of playground aggression at pretest (Kellam et al., 1998). Also, the Anger Coping Power evaluation (Lochman et al., 2001) reported stronger intervention effects on boys' aggression for those who reported lower baseline scores on problem solving and higher baseline scores on anxiety and somatization. In contrast, other program evaluations have found somewhat better results for low-risk compared with high-risk participants (Perry et al., 2002). For example, significant intervention effects on past-month smoking and lifetime marijuana use were found for SFA students who had not initiated drug use at baseline, but these effects were nonsignificant for baseline substance users (Eisen et al., 2002). An evaluation of Project Alert (Ellickson & Bell, 1990) demonstrated that cigarette "experimenters" (those who had initiated use, but smoked less than three times per year) who received the school-based substance prevention curriculum from either teachers or teachers assisted by teens were less likely than control students to report past-week or past-month smoking after the intervention ended. Current cigarette users who received the Alert curriculum from teachers assisted by teens, however, reported higher rates of past-week and past-month smoking compared with control students. These studies indicated that interventions may differentially affect high- and low-risk students, though the direction and magnitude of such effects may be program specific.

These data suggest that it is important not only to test for main effects of preventive interventions but also to evaluate potential variations in effects for different groups of participants. In this article, we propose the use of multilevel mixture models to examine the effects of a community-based prevention program on participants with different profiles of problem behaviors and delinquency. The use of these models is demonstrated using data from the Community Youth Development Study (CYDS).

THE COMMUNITY YOUTH DEVELOPMENT STUDY

CYDS is a group randomized trial testing the efficacy of the Communities That Care (CTC) prevention operating system (Hawkins & Catalano, 1992; Hawkins, Catalano, & Arthur, 2002). CTC is a community-based, strategic approach designed to reduce youth involvement in problem behaviors, including substance use, delinquency, and violence. Community members are trained to assess the epidemiology of problem behaviors in their community, identify risk and protective factors that influence the likelihood of these behaviors, prioritize specific factors as targets for intervention, and address these factors with evidence-based prevention programs.

Study Aims

CYDS has several design features that guided this demonstration of finite mixture modeling. First, it evaluates a universal prevention operating system that, targeting all youth in the community, aims to reduce the full range of substance use and problem behaviors via the implementation of selected research-based programs (Hawkins et al., in press). The first aim of the present study is to implement mixture models to identify groups of students mostly likely to be affected by the CTC intervention. Second, because the CYDS is a group randomized trial, multilevel analyses are needed in order to model the intervention effects at the community level. The second aim of this study is to demonstrate how the multilevel data structure can be taken into account with random-effects mixture models. Third, the logic for identifying latent classes of students in this study implies that these classes represent qualitatively different groups of students. A third aim of this study is to establish the validity of these classes by examining their relationship with exogenous variables (e.g., risk and protective factors) and by replicating the results for two cohorts of students. The fourth aim of this study is to examine differences in levels of behavior problems between students in treatment and control communities at the end of the community mobilization and planning phase of the CTC intervention in the CYDS. Because the 1st year of the intervention involved prevention training and planning but not the implementation of preventive interventions, which could be expected to affect behaviors, we expect no treatment effects at this point. However, the analyses can nonetheless illustrate the use of multilevel mixtures to test intervention effects that might occur later in later years of the CYDS. This study also examines power to find intervention effects in the CYDS. The power analyses support the utility of using multilevel mixtures to examine intervention effects and are used to guide the model specification for the intervention test.

MIXTURE MODELS IN INTERVENTION RESEARCH

A mixture model is a statistical tool for identifying latent groups of individuals (McLachlan & Peel, 2000; B. O. Muthén & Shedden, 1999). The mixture model assesses unobserved variables on a categorical scale with the goal of identifying latent classes of respondents. Mixture models are starting to be used to evaluate whether interventions have different effects for different types of students (B. O. Muthén et al., 2002). For example, an evaluation of the Good Behavior Game examined whether the effects of the intervention were different for students with differing trajectories of aggressive behavior (Stoolmiller et al., 2000). The evaluation hypothesized that students with moderate levels of aggression would be most affected because the level of the intervention delivered was not expected to be strong enough to affect students with the highest levels of aggressiveness (B. O. Muthén et al., 2002).

Most previous work has used mixture models to examine differential intervention effects with repeated measures by identifying latent classes of individuals with differing trajectories and then assessing within class differences in the effects of the intervention on development (B. O. Muthén et al., 2002). In the present study we look at intervention effects between classes rather than within classes. We ask whether the intervention affects the proportion of students in a given class rather than whether the effect of the intervention differs for students in different classes. When there are more than two latent classes, this may be thought of as a differential effect of the intervention because students in some classes may be affected while those in other classes are not affected. It can also be viewed as a nonlinear intervention effect because the intervention is modeled as affecting students with particular profiles of the outcome variables. This question is particularly relevant for preventive interventions where it is believed that the intervention may reduce one type of problem behavior but not others. For example, a population-based substance use intervention that reaches a large proportion of the population with a relatively low dosage might be expected to reduce experimental drug use but not more serious levels of use.

Figure 1 depicts the proposed model for examining effects in the CYDS. The model includes a latent class variable identified by 12 indicators of substance use and delinquency. As CYDS is a group randomized trial, intervention effects are modeled as occurring at the community level. Effects of CYDS on class membership are tested because the theory guiding the present analyses suggests that CTC affects the probability of class membership. However, a variety of different models looking at intervention effects could be incorporated into the multilevel mixture framework.

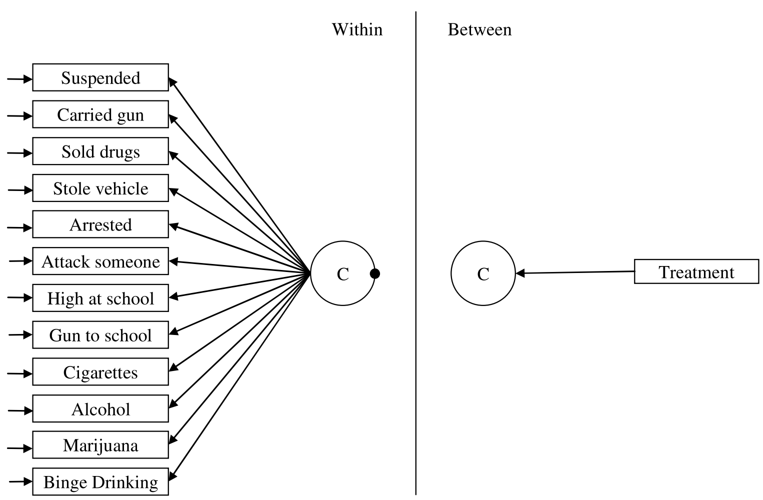

FIGURE 1 Multilevel mixture model for the effects of treatment on class membership in CYDS.

ASSESSING VALIDITY OF THE LATENT CLASSES

To better understand the latent classes identified in these analyses, we used procedures from the field of psychometrics that emphasize the assessment of validity when measuring latent constructs (Clark & Watson, 1995; Cronbach & Meehl, 1955). We propose that when latent class variables are used to identify qualitatively distinct classes of individuals they be thought of as any other latent variable and subjected to a similar process of measure development and validation. Ideally this process should be theory driven, utilizing previous knowledge of what types of qualitative differences might exist and what variables might predict those differences. The measurement model for a latent class variable consists of two parts: the number of classes and the item distributions/probabilities within each class. This is similar to exploratory factor analysis, in which the user specifies the number of factors and then the optimal weights of each item on those factors are derived analytically. Unfortunately, as in exploratory factor analyses, the results are sample dependent. We propose that one important step in providing evidence for the validity of latent classes is that the analysis be replicated in an independent sample. This should serve to reduce the chance that results are due to sample-specific characteristics. Replication in a separate cohort also provides some evidence that the results are not specific to a particular cohort.

Another step in establishing the validity of latent classes is to provide evidence for construct validity. If classes are identified based on theoretical and empirical evidence of what is known about the construct under study, then theory can be used to support construct validity. One way to provide evidence for validity, drawing from Cronbach and Meehl's (1955) use of the nomological net, is to show that the classes relate to other variables differently and in ways predicted by theory. In the CYDS example, one way to establish validity is to show that different variables (i.e., risk factors) predict membership in different classes defined by students' self-reported levels of involvement in problem behaviors. If three classes are identified in the CYDS analyses (e.g., Abstainers, Experimenters, and Drug Users) and theory or previous evidence shows that Experimenters tend to be adolescents who experienced poor family management practices, whereas Drug Users tend to be those whose friends engage in substance use, then partial evidence for the validity of the classes will be that these variables differentiate the groups of students.

Although latent classes may represent qualitative differences between participants, an alternative hypothesis is that the classes represent simple variation on a continuum of problem behaviors and accordingly are rank ordered on the predictor variables. This alternative has implications for interpreting findings of differential intervention effects. If true, this would represent a nonlinear intervention effect rather than suggesting that the intervention affects some groups differently than others. In this investigation, based on previous theoretical and empirical work suggesting different developmental pathways for substance use and delinquency (Moffitt, 1993), we expect that analyses will identify qualitative differences among participants. We test this assumption by replicating the results of the initial mixture analysis in a second data set and examining evidence for the construct validity of the classes identified.

GROUP RANDOMIZED TRIALS

This study examines the use of mixture models within a group randomized trial (GRT). GRTs involve groups or clusters of individuals randomized into intervention and control conditions (Feng, Diehr, Paterson, & McLerran, 2001; Feng et al., 1999; Murray, 1998; Raudenbush & Liu, 2000). In prevention and biomedical research, this design is useful when interventions are provided to groups or to all individuals within an organization. Because groups, not individuals, are randomly assigned to conditions in GRTs, intervention must be thought of as a group-level rather than an individual-level variable. Typically, individuals within groups are more similar to each other than to those in other groups, a condition that leads to a violation of the assumption of independence of observations made by fixed-effects statistical models (e.g., regression type models). When fixed-effects models are used with this design, standard errors will be biased downward (Murray, Clark, & Wagenaar, 2000; Murray, Feldman, & McGovern, 2000). A second challenge associated with the use of GRTs is that the appropriate degrees of freedom for the intervention effect are based on the number of groups randomized rather than on the number of individuals in the study (Murray, 1998). Therefore, when analyzing data from GRTs, it is recommended to use random-effects models that assume individuals are independent of each other conditional on group membership and to estimate the correct degrees of freedom for group-level effects (Murray, 1998; Raudenbush & Bryk, 2002).

The use of random-effects models in a regression context is common and has been well documented (Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). The inclusion of random effects in the estimation of mixture models is less developed, though Vermunt (2003) has described a mathematical model and addressed issues with the specification and use of random-effects mixture models. The current article applies multilevel mixture models to group randomized trials. The illustration of these techniques with CYDS involves the specification of both fixed- and random-effects portions of the model, examination of conditions necessary for identification and successful estimation of the model, and methods for assessment of power to detect intervention effects.

The model tested in this study (see Figure 1) explicitly accounts for the fact that students in the demonstration data set were nested within schools and within communities. The regression weights of each latent class on the items measuring substance use and delinquency (in the within section of Figure 1) were constrained to be the same across clusters. The model included one latent class variable with T classes so that each student had a nonzero probability of being in each of the T classes. At the school level, the latent class variable represents the proportion of students in each school who were estimated to be in each of the latent classes. The regression of the latent class variable on intervention status at the community level indicates that community involvement in the intervention can affect the proportion of students in each class, thus answering the research question, "Does the intervention affect the probability that a student will be in a particular class as compared with a reference class?" T-1 parameters were estimated for this effect, as the latent class variable has T-1 degrees of freedom. The solid dot at the individual level in Figure 1 illustrates that the probability of membership in a given class is allowed to differ among schools; this is the random-effects portion of the model.

THE FINITE MIXTURE MODEL

Here we introduce the mathematical formulation of the finite mixture model that was implemented in this article. The formulation presented here is for ordered categorical indicator variables. Thus, in this case the finite mixture model simplifies to a latent class model, which follows. Let the K-dimensional vector of unordered responses be Y_1 through Y_K for individual i within group j be denoted by $\mathbf{Y}_{ij}$ with $i = 1, \ldots, n_j$; $j = 1, \ldots, J$. Let c_{ij} denote the unobserved latent class variable and let a particular class be $t = 1, \ldots, T$.

The simple latent class model can then be written as

$$\begin{aligned} P(\mathbf{Y}_{ij} = \mathbf{s}) &= \sum_{t=1}^{T} P(c_{ij} = t) P(\mathbf{Y}_{ij} = \mathbf{s} | c_{ij} = t) \\ &= \sum_{t=1}^{T} P(c_{ij} = t) \prod_{k=1}^{K} P(Y_{ijk} = s_k | c_{ij} = t), \end{aligned} \tag{1}$$

where $\mathbf{s}$ is a specific response pattern and s_k is the response for the k^{th} item. It models the probability of a particular response as the weighted average of the probabilities conditional on class membership. The weights, $P(c_{ij} = t)$, are the probability of individual i belonging to class j. Notice that the items are assumed to be independent given membership in a specific latent class. To account for the multilevel structure of the model, the parameters are allowed to depend on the group j. Because the particular interest of this study is in the weights as well as the probabilities conditional on class membership, we can write them using the logit function, $\ln\left(\frac{p}{1-p}\right)$, which when applied to Equation (1) leads to

$$P(c_{ij} = t) = \frac{\exp(\gamma_{tj})}{\sum_{r=1}^{T} \exp(\gamma_{rj})} \tag{2}$$

$$P(Y_{ijk} \geq s_k | c_{ij} = t) = \frac{exp(\beta^k_{s_k t_j})}{\sum_{r=1}^{S_k} exp(\beta^k_{r t_j})}, \tag{3}$$

where S_k is the number of response categories, which are assumed to be ordered. Equation (2) in this model corresponds to the probability of individual i in group j being in class t, while Equation (3) relates the response vector $\mathbf{Y}$ to each class. For identification purposes the coefficient in the reference category needs to be fixed at zero. This formulation is equivalent to fitting separate models for each group j. Because of the complexity and large number of parameters in this model, the following more restrictive model, for which the parameters are not allowed to depend on a specific group, is often more useful. This is called the fixed-effects model in this study and is the starting point for our analyses.

$$P(c_{ij} = t) = \frac{exp(\gamma_t)}{\sum_{r=1}^{T} exp(\gamma_r)} \tag{4}$$

$$P(Y_{ijk} \geq s_k | c_{ij} = t) = \frac{exp(\beta^k_{s_k t})}{\sum_{r=1}^{S_k} exp(\beta^k_{rt})}. \tag{5}$$

Notice that, again, identifying constraints like $\gamma_1 = \beta^k_{1t} = 0$ have to be placed on the parameters.

Random Effects in the Finite Mixture Model

The fixed-effects model assumes that the proportion of participants in each class is the same across all clusters. An alternative to using fixed effects for the probability of class membership is a model in which the class-specific effects are assumed to be random from a particular distribution. To do so, the general multiclass model from the previous section, or more specifically the coefficients in Equation (2), could be defined as

$$\gamma_{tj} = \gamma_t + \tau_t \cdot z_j \, \forall t = 1, \ldots, T, \tag{6}$$

with $z_j \sim N(0, 1)$, and $\gamma_1 = \tau_1 = 0$ for identification. The variance of z_j may also be estimated as in $z_j \sim N(0, \sigma)$, with one of the τ_t parameters fixed to 1 for identification. The coefficient γ_{tj} therefore assumes that the between-group variation in the log-odds of belonging to the t^{th} instead of the first latent class follows a normal distribution with mean γ_t and a variance of $\tau_t \cdot z_j$. Notice that a perfect correlation between the random components in the γ's is implied. This is true because the same random effect z_j is used for each class t and then scaled by τ_t. In this study, this is called the random factor model because it can be expressed by having one factor with a random variance for the between-school differences in the probabilities of class membership. The rather restrictive assumption that γ's are perfectly correlated can be relaxed by using a covariance structure with a (T-1)-dimensional distribution on the γ's. Instead of using (T-1) perfectly correlated normal random variables, the (T-1)-dimensional vector $\boldsymbol{\gamma}$ can be modeled with a multivariate normal distribution. That is, $\boldsymbol{\gamma} \sim MN(\boldsymbol{\mu}, \sum)$, where $\boldsymbol{\mu}$ represents a vector of means and $\boldsymbol{\Sigma}$ is the variance-covariance matrix. In this study, we call this the fully random model.

Covariates in the Finite Mixture Model (Combining Individual Covariates and Group-Level Effects)

Let $\mathbf{X}_{1j}$ denote the r_J-dimensional vector of covariates of the group j and let $\mathbf{X}_{2ij}$ be the r_I-dimensional vector of covariates for the individual *i*. Define γ_{0tj} as the class intercept, $\boldsymbol{\gamma}_{\mathbf{1t}}$ as the r_J-dimensional vectors coefficients in the classes, and $\boldsymbol{\gamma}_{\mathbf{2t}}$ as the r_I-dimensional vector of coefficients in the individual. A multinomial logistic regression model for c_{ij} is obtained by

$$P(c_{ij} = t | X_{1j}, X_{2ij}) = \frac{exp\left(\gamma_{0tj} + \boldsymbol{\gamma}'_{\mathbf{1t}}\mathbf{X_{1j}} + \boldsymbol{\gamma}'_{\mathbf{2t}}\mathbf{X_{2ij}}\right)}{\sum_{r=1}^{T} exp\left(\gamma_{0rj} + \boldsymbol{\gamma}'_{\mathbf{1r}}\mathbf{X_{1j}} + \boldsymbol{\gamma}'_{\mathbf{2r}}\mathbf{X_{2ij}}\right)}. \tag{7}$$

Notice that, depending on the definition of the parameters γ_{0tj}, $\boldsymbol{\gamma}_{\mathbf{1t}}$ and $\boldsymbol{\gamma}_{\mathbf{2t}}$, this formulation can be used to deal with fixed effects as well as random effects.

Interpretation of the Model Parameters

Figure 1 presents the graphical depiction of this model applied to the CYDS.

The within-effects of Figure 1 correspond to covariates $\boldsymbol{\gamma}_{\mathbf{2t}}, \boldsymbol{\gamma}_{\mathbf{1t}}$, while the between-effect relates to the intervention effect γ_{0tj}. The estimates of β from Equation (5) correspond to the arrows from the latent class variable to each of the K indicators.

Parameter estimation. To estimate the parameters of the model, the maximum likelihood approach is used. Because of the large number of parameters, the necessary maximization is rather expensive computationally and, therefore, not straightforward. Instead of using a Quasi-Newton method on the complete data likelihood, a modified EM-Algorithm is used.

First, the expected value is approximated by replacing the integral with a finite sum of points, $m = 1, \ldots, M$. Using the conditional independence assumption of the classes, then $P(W_j = m, c_{ij} = t|Y_j, X_j)$, where W_j represents the probability mass for group j at a certain point. This can be computed directly without having to find the much more complicated $P(W_j = m, c_j = t|Y_j, X_j)$ first. Notice that this general formulation allows W_j to come from any mixing distribution, not only the normal mixing distribution used in this article. A more detailed description of the modified EM-Algorithm can be found in B. O. Muthén & Shedden (1999) and Vermunt (2003).

Understanding the model. The model described in this article is fairly complicated, requiring the estimation of many parameters and having many alternative specifications that may affect the results. A reasonable concern for models with this degree of complexity is that the analyses are too far removed from the data, making it difficult to understand and have confidence in the results. Because the models described here have response profiles to multiple items as their outcome, it is not possible to translate these results to simple mean differences on any item. However, researchers familiar with basic latent class analysis (see McLachlan & Peel, 2000) may benefit by thinking of these models as similar to a two-stage approach. In the first stage, separate latent class analyses are conducted for each community with the substance use and delinquency variables as the indicators of the latent classes. Individuals are then assigned to their most probable latent class. In the second stage, the proportion of students in each community in each class is regressed on the treatment effect. In essence, these models work by identifying students as belonging to a particular class and then examining the relationship between those classes and treatment. Although this technique oversimplifies the analyses, it may help in understanding what is being tested and perhaps provide a technique for assessing some of the assumptions of the model, such as the assumption that all communities contain the same latent classes.

METHOD

Analyses for the first three aims of this article used data obtained from statewide surveys of public school students conducted by the state agencies responsible for alcohol and drug abuse prevention in the states of Colorado, Illinois, Kansas, Maine, Oregon, Utah, and Washington. These states used one of two methods for drawing statewide samples of schools for their surveys. In Kansas and Washington, all public schools in the state were invited to participate in the surveys and schools that elected to participate were included in the samples. The other five states used a probability sampling scheme stratified by the geographic regions of the state to draw representative samples of schools.

The fourth aim uses data from the CYDS trial, which was collected from 24 communities (12 matched pairs[1]) in the same seven states. This data was collected by CYDS staff in all states except Utah and Kansas, where it was collected by the same agencies as in the first samples. None of these 24 communities were included in the first two data sets. For the CYDS trial, one community from each matched pair was assigned randomly to be in the intervention condition and implement CTC and the other to be in the control condition with prevention services operating as usual.

Surveys were administered during one classroom period using standardized administration procedures. Administration procedures ensured the anonymity and confidentiality of students' responses. Screening criteria were used to exclude respondents who lied or responded inconsistently to a set of screening items. For instance, students were asked if they had ever smoked cigarettes in their lifetime and also how frequently they had smoked cigarettes during the past 30 days. If students responded "Never" to the question, "Have you ever smoked cigarettes?" and also indicated that they smoked one or more cigarettes per day during the past 30 days, their answers were considered inconsistent. Data used for model development were collected in 1998 and 2000 from a sample of schools that excluded schools in the 24 CYDS communities; data used to test intervention effects in the CYDS study were collected in 2004. The 1998 data set was used in an exploratory fashion to develop the measurement model and the 2000 data were used to replicate these results. It is especially important to replicate these results in a separate cohort of students to reduce the chances that the results are cohort specific.

---

[1]Matched pairs were used in this study to increase the likelihood of baseline comparability. However, explicitly including the matching in analysis when there are a small number of clusters causes a reduction in power because of a loss of cluster level degrees of freedom. With low ICCs for the outcomes it is advisable to ignore the matching when conducting analyses (Diehr, Martin, Koepsell, & Cheadle, 1995).

Participants

The 1998 exploratory data set included 44,531 eighth-grade students from seven states, nested within 494 schools or, in one state, counties.[2] The second data set, used for confirmatory analyses, included data from 45,848 eighth-grade students in four states in 2000 nested within 418 schools. The 2004 data set used to test intervention effects was comprised of eighth-grade students from the communities in the CYDS study (Hawkins et al., in press) and was collected at the end of the mobilization and planning phase of the CYDS intervention. Because the CTC intervention was in the planning phase in the 1st year, few prevention activities were conducted that could reasonably be expected to influence student behaviors. Consequently, these analyses serve primarily to assess baseline comparability of the treatment and control conditions. No intervention effects are hypothesized.

Measures

Data for all analyses were collected using the Communities That Care (CTC) Youth Survey (Arthur, Hawkins, Pollard, Catalano, & Baglioni, 2002; Glaser, Van Horn, Arthur, Hawkins, & Catalano, 2005), which was administered anonymously by teachers during one classroom period. The CTC Youth Survey measures risk and protective factors for problem behaviors as well as self-reported substance use and delinquency and students' demographic characteristics. The current analyses focused on the substance use and delinquency items. Students were asked to report the frequency with which they used alcohol, tobacco, and marijuana in the last 30 days and in their lifetime as well as the number of times in the last 2 weeks they had more than five drinks at one time (binge drinking). Lifetime and 30-day use for each of these outcomes were combined into one variable with three levels: no use, lifetime but not current use, and current use within the last 30 days. These responses are treated as ordered categories. Students also indicated the frequency with which they engaged in eight delinquent behaviors over the last year. Each of the eight delinquency items and the binge drinking question were treated as binary variables. Frequency distributions for each of the substance use and delinquency variables for the 1998 and 2000 data are reported in Table 6.

Two variables, lifetime alcohol and marijuana use, were not included in one state in the 2000 sample. However, two proxy variables, age at first use

[2]In one state, counties rather than schools were used because school identifiers were not available. However, in most cases only one school per county was sampled. An examination of numbers of individuals per cluster indicates that the counties had approximately the same number of students as schools in other states except for two counties with over 1,000 students, which were dropped from the analysis.

of alcohol and marijuana, were included. To avoid excluding data from an entire state or using full information maximum likelihood (FIML) estimation of missing data for these two variables, scores for the lifetime use variables were imputed using single imputation based on the relationship of the age of first use variables with lifetime use in the other states. A high degree of correspondence between the proxy variables and the lifetime use variables was observed for the states for which both variables were present, supporting the validity of this method.

Data Analysis

Data analyses were conducted in Mplus Version 3.12 (L. K. Muthén & Muthén, 2004) and Version 4.2 (L. K. Muthén & Muthén, 2006) for the analyses of CYDS data. For missing data due to students skipping questions, the full information maximum likelihood (FIML) estimator was used. FIML uses the observed data under the assumption that missing data points are missing at random.

RESULTS

Specification of Fixed Effects

Analyses based on the exploratory data set began with the specification of the fixed-effects portion of the mixture model (i.e., determining the number of latent classes to be estimated and the interpretation of those classes). Although the random-effects portion of the model may change the specification of the fixed-effects portion, we began by specifying the fixed effects because these were easier to estimate and provided guidance for the random-effects models. The initial model had two levels. The individual level included the latent class measurement model, and the school level corresponded to school means and variances for each latent class. The analyses included 12 indicator variables at the individual level and one latent class variable with T classes (see Figure 1). Because the indicator variables were binary or ordinal, these patterns were captured by thresholds estimated for T-1 classes. For the binary responses, the thresholds represent the log-odds of responding “no” versus “yes” for an individual in class t.

One further issue with model estimation is noteworthy. Finite mixture models tend to be highly sensitive to the choice of start values (Bauer & Curran, 2004; McLachlan & Peel, 2000). In this study, 100 different random start values were used for each fixed-effects model. Each set of start values was run for 10 iterations, and the models with the 10 best log likelihoods were continued until the model converged. Differences in the log likelihoods among the top models were less than .01, providing evidence that the solution is not a local maxima. Further, following recent simulation analyses looking at the effects of local maxima (Hipp & Bauer, 2006), the final results were reanalyzed with 1,000 random starts, each run until convergence. In some cases results from analyses that converged to different log likelihoods were also examined; no substantive differences in conclusions were found.

Analyses were conducted for models with one through nine latent classes. A common method for choosing the number of classes is to use penalized information criteria such as the Akaike Information Criteria (AIC; Akaike, 1974) and the Sample Size Adjusted Bayesian Information Criteria (SBIC; Sclove, 1987). These values are reported in the five rows of Table 1 for fixed-effects models with one through six classes. The traditional criterion is that the best-fitting model is the one with the lowest value on the penalized information criteria. However, in this study, it is clear that this criterion was never met, likely because the sample size increased the power of these analyses to find latent classes that were either very small or were due to correlated errors rather than "true" categorical differences among groups of individuals. Another test

TABLE 1
Fit Criteria for Different Model Specifications

	Number of Classes					
Model	*1-Class*	*2-Class*	*3-Class*	*4-Class*	*5-Class*	*6-Class*
Fixed effect model						
# of Parameters	15	31	47	63	79	95
Loglikelihood	−237012	−193045	−185024	−182245	−180734	−179737
AIC	474113	386151	370141	364615	361626	359663
SBIC	474196	386322	370401	364963	362062	360187
Entropy	1	0.895	0.812	0.802	0.801	0.802
Random class model						
# of Parameters	N/A	32	49	66	83	**
Loglikelihood		−192539	−184359	−181699	−180122	
AIC		385143	368817	363529	360410	
SBIC		385319	369087	363893	360868	
Entropy		0.895	0.813	0.803	0.785	
Random Factor						
# of Parameters	N/A	N/A	49	66	83	100
Loglikelihood			−184370	−181574	−180419	−179040
AIC			368838	363279	361005	358280
SBIC			369109	363643	361463	358831
Entropy			0.812	0.803	0.789	0.800

Note: N/A–not applicable, ** this model is computationally too difficult to run.

for empirically determining the number of latent classes is the bootstrapped likelihood ratio test (BLRT; McLachlan & Peel, 2000). This test, however, is computationally intensive and, in this analysis, is also affected by the increased power due to the large sample size. To establish that the BLRT finds too many classes with this sample size, we estimated it for the six-class model, which we rejected as described later in this article because of a small and uninterpretable class, and found that it did, indeed, indicate the need for six rather than five classes. Thus, in selecting the number of classes to be retained, we relied partially on penalized information criteria and changes in log likelihood values to eliminate those solutions with very poor results, similar to using a scree test of these values (Nylund, Asparauhov, & Muthén, 2007). Large reductions in these fit indices were taken to indicate that additional classes were needed. We also looked at the substantive interpretation of the different classes.

In these analyses, classes were interpreted based on the responses of students in a given latent class to the indicators of the latent class variable (i.e., substance use and delinquency items). Rather than present the thresholds (i.e., the parameters actually estimated), which are difficult to interpret, Table 2 presents the probability of a given response to each item for students in each latent class. For brevity, we focus on four-class and five-class models, although one- through nine-class models were examined. Solutions with fewer than three classes were eliminated because they were clearly inadequate to fit the data, and models with more than five classes were eliminated because those solutions subdivided the simpler solutions into very small classes (containing less than 3% of the students) and had little practical or theoretical value. Note that the latent class model described assumes that the indicators are independent given class membership. In this example there are some cases for which this could be questioned, for example, whether marijuana use and getting high at school are really independent within classes. This possibility was examined by looking at within-class bivariate residuals. The highest residuals found were .05 and nearly all were under .01. No evidence was found suggesting that additional within-class correlations between variables were needed.

Both four- and five-class solutions had a large latent class containing about half the sample who did not engage in any problem behaviors, with the exception of having a small probability of reporting lifetime cigarette and alcohol use. We termed these students "Abstainers." Each of these solutions also had a group of students, comprised of just over 10% of the sample, who engaged in a high level of substance use but who had a fairly low probability of reporting other problem behaviors except *getting high at school* and *attacking someone*. We referred to these students as "Drug Users." Each class also had a group of students comprising 4% to 5% of the sample whose probability of engaging in substance use was as high or almost as high as the Drug Users but who were differentiated by their increased probability of engaging in problem

TABLE 2
Results of Multilevel Unconditional Latent Class Analyses for Eighth Graders

	4-Class				*5-Class*				
	1	2	3	4	1	2	3	4	5
Proportion	50%	31%	14%	5%	52%	23%	11%	11%	4%
Cigarettes									
Never	0.90	0.20	0.04	0.10	0.90	0.12	0.21	0.02	0.12
Lifetime	0.09	0.65	0.31	0.25	0.09	0.74	0.42	0.30	0.25
Current	0.01	0.16	0.64	0.65	0.01	0.14	0.37	0.68	0.63
Alcohol									
Never	0.73	0.10	0.01	0.06	0.71	0.13	0.01	0.01	0.07
Lifetime	0.22	0.54	0.12	0.15	0.23	0.69	0.07	0.14	0.16
Current	0.05	0.36	0.86	0.79	0.06	0.19	0.92	0.85	0.78
Marijuana									
Never	1.00	0.74	0.16	0.13	1.00	0.70	0.64	0.06	0.16
Lifetime	0.00	0.22	0.24	0.16	0.00	0.25	0.18	0.23	0.16
Current	0.00	0.05	0.60	0.71	0.00	0.05	0.17	0.71	0.68
Binge Drinking									
No	1.00	0.86	0.33	0.27	1.00	0.98	0.41	0.33	0.28
Yes	0.00	0.14	0.68	0.73	0.00	0.02	0.59	0.67	0.72
Suspended									
No	0.96	0.86	0.66	0.23	0.96	0.83	0.88	0.56	0.23
Yes	0.04	0.15	0.34	0.77	0.04	0.17	0.12	0.44	0.78
Carried gun									
No	0.98	0.95	0.91	0.28	0.98	0.95	0.94	0.89	0.16
Yes	0.02	0.05	0.09	0.73	0.02	0.05	0.06	0.11	0.84
Sold drugs									
No	1.00	0.99	0.79	0.23	1.00	0.99	1.00	0.67	0.21
Yes	0.00	0.01	0.22	0.77	0.00	0.01	0.00	0.34	0.79
Stole Vehicle									
No	1.00	0.99	0.92	0.40	1.00	0.98	0.98	0.89	0.33
Yes	0.00	0.01	0.08	0.60	0.00	0.02	0.02	0.12	0.67
Arrested									
No	0.99	0.96	0.81	0.26	0.99	0.94	0.97	0.71	0.24
Yes	0.01	0.05	0.19	0.74	0.01	0.06	0.03	0.29	0.76
Attacked someone									
No	0.96	0.84	0.60	0.12	0.96	0.83	0.81	0.51	0.10
Yes	0.04	0.16	0.40	0.88	0.04	0.17	0.19	0.49	0.90
High at school									
No	1.00	0.95	0.39	0.14	1.00	0.95	0.87	0.24	0.16
Yes	0.00	0.05	0.61	0.86	0.00	0.05	0.13	0.76	0.85
Gun to school									
No	1.00	1.00	0.99	0.57	1.00	1.00	1.00	0.99	0.76
Yes	0.00	0.00	0.01	0.43	0.00	0.00	0.00	0.01	0.54

Note: Proportions are based on estimated posterior probabilities.

behaviors; for example, the probability of these students carrying a gun to school in the last year was between 43% and 54%. We termed these “Problem Students.”

The difference between the four- and five-class models was primarily with a class of students we called “Experimenters.” In the four-class model, this group comprised 31% of the sample. Experimenters have high lifetime alcohol and cigarette use but low current use. In the five-class model, this group comprised 23% of the sample, and there was a second class containing 11% of the sample. Based on a cross tabulation comparing class membership in the four- and five-class models, the new class was comprised of students from both the Experimenter and Substance User classes from the four-class model. In the five-class model, these students had very high current alcohol use but low rates of all other behaviors except lifetime cigarette use. We termed this the “Alcohol Use” class. The evidence thus far provided support for either the four- or five-class solutions. Further analyses focused on the evaluation and comparison of these models.

Specification of the Random Effects

The difference between the fixed- and random-effects models involved the estimation of the variance of the latent class means. The models described earlier had class means that were estimated for T-1 classes. The class means represented the log-odds of being in a given class versus being in the reference class (the Abstainer class in this application), although when random effects are included in the model they are only estimates of the log-odds. In the fixed-effects models, the variance of the class means between schools was fixed to zero whereas in the random-effects model, a variance was estimated for these parameters. Fixing this variance to zero amounted to making the assumption that the proportion of students in each class was the same for each school. Although it is beyond the scope of this article to evaluate the effects of violating this assumption, we suspect that they would be the same as in ordinary regression analysis. That is, when conducting a group randomized trial, the use of a fixed-effects model can lead to underestimation of standard errors and, consequently, an increase in Type I errors. Thus, our next step was to model and quantify the between-school variation in class membership.

Alternative specifications of the random-effects models were considered (see Figure 2 for graphical depictions of the options considered). In the first specification, each of the T-1 class means was allowed to vary among schools with covariances among these effects freely estimated. Thus, every school may have had a different proportion of students in every class. Although this was a compelling model, computationally it became increasingly difficult to estimate as the number

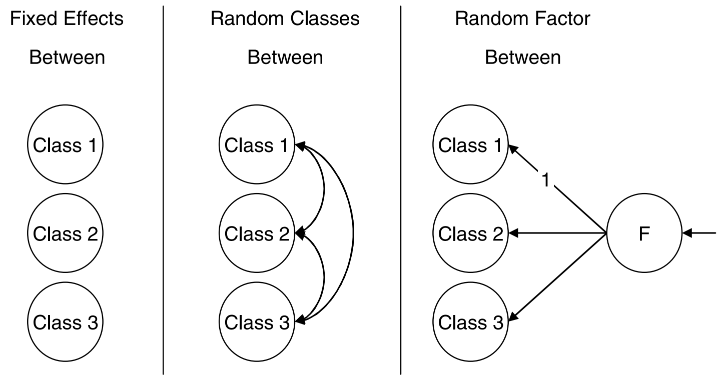

FIGURE 2 Alternative specifications of the between schools random effects.

of classes increased.[3] Vermunt (2003) suggested an alternative specification that takes advantage of the fact that the variances of the means for each class are frequently highly correlated. He suggested the use of a common factor to capture these covariances and the inclusion of the random effect by allowing the variance of this factor to vary across clusters (see the random factor model in Figure 2). This specification greatly simplifies model estimation, although as more classes are added it becomes increasingly restrictive. Hybrid models that combine these two specifications are also possible. In this study, three models were compared: the fixed-effects model, the fully random model, and the random-factor model. It should be noted that the fully random model was not estimable when more than five classes were included due to the computational complexity of estimating random effects.

Table 1 presents the log likelihood and penalized information criteria for each model. Note that although these models are nested, a chi-square difference test is not appropriate because the variances in the fixed-effects model are constrained to zero, which is on the boundary of the parameter space for the variance. However, comparison of model fit between the fixed-effects and random-effects models shows large differences in all criteria, demonstrating the need for a random error variance for the class means. Comparing values of penalized information criteria between the fully random and random-factor models showed only small differences favoring the least restrictive (fully random) models. Most other

[3]As currently implemented in Mplus, these models required numerical integration. One dimension of integration is required at the within-school level and each random effect at the between-school level requires an additional dimension. The number of integration points used was, by default, 15. However, the five dimensions of integration in the five-class model required that this be reduced to five integration points to be run on a computer with two gigabytes of RAM available to the program. These models took over 24 hr to converge.

analyses focused on the random-factor model because it was computationally easier to estimate than the fully random model.

Throughout the evaluation of these models, it was important to consider whether the estimation of random effects changed the results from the fixed-effects models. The penalized information criteria for these models (presented in Table 1) did change, but they changed in a very systematic way such that model fit improved for the less restrictive models by about the same magnitude for each model. Further, the profiles of the probability of a student responding positively to each item for each latent class were evaluated for the different models. These profiles indicated only very small differences between either the proportion of students in a given class for the three random-effects specifications or in the proportions of students expected to respond to each item. Comparing Tables 2 and 7 demonstrates the differences between the fixed-effect and random-factor models for the five-class solution. This illustrates that the fixed-effects portion of the model was essentially unchanged by the specification of the random effects. It should be noted that multiple starting values were used for the random-factor model and for some of the fully random models. However, in all cases we found that it was adequate to use the parameter estimates from the fixed-effects models as start values for the fully random models rather than using multiple starts.

The final step in the estimation of the random effects was to quantify the differences in the proportion of students in a given latent class between schools. This task was more straightforward with the fully random model; consequently, the results reported here are from that model. Table 3 shows the between-school variances for each model as well as intraclass correlation coefficients (ICCs). It also shows the expected odds of a student being in a certain class versus being in the reference class for a school at the 10th and 90th percentiles. Variances presented here are on the log-odds scale and the ICC was computed using the formula proposed by Hedeker (2003) and recommended by Vermunt (2003). The odds for schools at the 10th and 90th percentiles were calculated based on the

TABLE 3
School Level Variance in Class Membership

	4-Class				*5-Class*				
	1	2	3	4	1	2	3	4	5
Proportion	50%	32%	14%	5%	51%	23%	11%	11%	4%
Average log-odds	ref	−0.51	−1.466	−2.772	ref	−0.87	−1.63	−1.78	−2.99
Variance in log-odds	ref	0.14	0.39	0.77	ref	0.18	0.22	0.50	0.75
ICC	ref	0.04	0.11	0.19	ref	0.05	0.06	0.13	0.19
School at 10th %	ref	0.27	0.09	0.02	ref	0.20	0.10	0.06	0.02
School at 90th %	ref	0.49	0.34	0.16	ref	0.42	0.26	0.29	0.13

assumption that the variance components follow a normal distribution. To find the expected odds of a student being in a given class for a school at the 10th percentile, we multiplied the square root of the variance (the standard deviation) in the log-odds by 1.28 (the z score for the 10th percentile) and subtracted the result from the average log-odds. The resulting log-odds was then converted to an odds. The expected odds for a school at the 90th percentile was found using the same formula except that 1.28 times the standard deviation was added to the average log-odds rather than subtracted from it.

All models used the Abstainer class (i.e., the largest class) as the reference category. The results indicated that the variance for the proportion of students in each latent class versus the Abstainer class differed greatly among classes. ICCs for membership in the Experimenter class versus the Abstainer class were relatively small (.04–.05) and similar to those reported for continuous variables in other studies (Hawkins, Van Horn, & Arthur, 2004). However, ICCs for the Problem Student class were quite large (.19), indicating that the odds of a student being in that latent class versus the Abstainer class differed greatly among schools.

Model Validation

The third step in the model development process involved understanding the differences among classes. We identified 13 risk factors as well as gender and ethnicity for which differences in the latent classes were expected. Using the exploratory sample, latent class membership was regressed on each one of these variables independently (see Table 4 for results from the four-class model and Table 5 for results from the five-class model). Independent models were run due to the relatively large degree of collinearity among risk factors. For the risk factors, the results represent the expected increase in the log-odds of being in each class versus the Abstainer class that is associated with a one-standard-deviation increase in the risk factor. The demographic variables were dummy coded so the parameter estimates can be interpreted as the difference in odds between that group and the reference group.

These analyses showed that these predictors differentiate some classes but not others. One example is that girls were less likely than boys to be in each class versus the Abstainer class, except that in the five-class model, girls were more likely than boys to be in the Alcohol Use class. This is one indication that the Alcohol use class was qualitatively different from the Experimenter and Drug Use classes. Another indication that latent classes were qualitatively different was that students in the Alcohol Use class had lower reports of school failure than all other classes, including the Experimenter class. There was also some support for qualitative differences between the Drug Use class and the Problem Student class. This was most clearly demonstrated with regard to the risk factors

TABLE 4
Predictors of Latent Class Membership with the Four-Class Model

	Experimenters			*Drug Users*			*Problem Students*		
	Beta	*SE*	*OR*	*Beta*	*SE*	*OR*	*Beta*	*SE*	*OR*
Demographics									
Female	**−0.15**	(.03)	1.09	0.09	(.05)	0.86	**−1.36**	(.07)	0.26
African American	**0.49**	(.10)	1.63	**0.33**	(.11)	1.40	**1.44**	(.12)	4.23
Native American	**0.63**	(.10)	1.89	**1.01**	(.12)	2.76	**1.59**	(.13)	4.88
Hispanic	**0.44**	(.06)	1.54	**0.72**	(.07)	2.06	**1.28**	(.09)	3.59
Asian	-0.09	(.07)	0.92	**−0.28**	(.10)	0.76	**0.61**	(.11)	1.83
Other	**0.24**	(.07)	1.28	**0.47**	(.09)	1.60	**1.11**	(.10)	3.03
Community Risks									
community disorganization	**0.50**	(.02)	1.66	**0.93**	(.03)	2.54	**1.46**	(.04)	4.30
norms favorable to drug use	**0.80**	(.03)	2.23	**1.38**	(.03)	3.97	**1.81**	(.04)	6.13
laws favorable to drug use	**0.69**	(.02)	1.99	**1.20**	(.03)	3.31	**1.60**	(.05)	4.94
perceived availability of drugs	**1.54**	(.03)	4.64	**2.96**	(.05)	19.20	**3.55**	(.10)	34.67
perceived availability of guns	**0.42**	(.02)	1.52	**0.71**	(.03)	2.03	**1.33**	(.04)	3.80
Family Risk									
parent attitude favorable to drugs	**2.46**	(.23)	11.69	**3.07**	(.24)	21.54	**3.49**	(.26)	32.92
parent attitude favorable to ASB	**1.24**	(.07)	3.44	**1.68**	(.07)	5.39	**2.12**	(.07)	8.31
School Risk									
academic failure	**0.68**	(.02)	1.97	**1.07**	(.03)	2.90	**1.52**	(.04)	4.58
Peer / Individual Risk									
favorable attitudes toward ASB	**1.71**	(.04)	5.51	**2.68**	(.05)	14.64	**3.64**	(.07)	38.21
favorable attitudes toward drugs	**3.74**	(.18)	42.27	**5.45**	(.21)	231.60	**6.21**	(.25)	495.71
perceived risk of drug use	**0.99**	(.04)	2.69	**1.77**	(.06)	5.85	**2.01**	(.08)	7.49
interaction with antisocial peers	**4.04**	(.19)	56.83	**5.22**	(.19)	184.56	**6.08**	(.20)	436.59
friends drug use	**4.27**	(.18)	71.24	**6.09**	(.19)	439.22	**6.94**	(.25)	1034.84
rewards for antisocial involvement	**0.73**	(.03)	2.07	**1.18**	(.04)	3.24	**1.58**	(.05)	4.84

Notes: Parameters indicated in bold are significat $p < .05$; OR is the Odds Ratio.

related to substance use. In most cases the regression weights for the Drug Use and Problem Student classes were quite similar.

There were, however, many other instances in which differences appeared to reflect that individuals were ranked from more to less behavior problems rather than that classes were qualitatively different. In most cases, the means of the risk factors increased from one class to the next, suggesting more quantitative than qualitative differences between classes. For example, each class was associated with a progressively higher level of the risk factor *Parental Attitudes Favorable to Drugs*.

Replication With a Second Data Set

The fourth step in these analyses was to replicate the results using the 2000 data set. We first examined differences in students' reports of engaging in problem behaviors between the two samples (see the first two columns in Table 6). In general, reports of problem behaviors declined from 1998 to 2000. Next, the random factor four- and five-class models were replicated. The change in

TABLE 5
Predictors of Latent Class Membership with the Five-Class Model

	Experimenters			*Alcohol Users*			*Drug Users*			*Problem Students*		
	Beta	*SE*	*OR*	*Beta*	*SE*	*OR*	*Beta*	*SE*	*OR*	*Beta*	*SE*	*OR*
Demographics												
Female	**−0.31**	(.05)	0.73	**0.27**	(.06)	1.31	-0.10	(.06)	0.90	**−1.51**	(.08)	0.22
African American	**0.78**	(.10)	2.18	**−0.58**	(.20)	0.56	**0.62**	(.11)	1.86	**1.53**	(.13)	4.64
Native American	**0.81**	(.11)	2.25	**0.25**	(.16)	1.29	**1.22**	(.12)	3.39	**1.58**	(.15)	4.83
Hispanic	**0.48**	(.07)	1.62	**0.41**	(.08)	1.51	**0.78**	(.08)	2.17	**1.33**	(.09)	3.78
Asian	**−0.03**	(.08)	0.97	**−0.26**	(.12)	0.77	**−0.21**	(.12)	0.81	**0.68**	(.12)	1.97
Other	**0.36**	(.09)	1.43	-0.05	(.12)	0.95	**0.61**	(.09)	1.84	**1.16**	(.12)	3.19
Community Risks												
community disorganization	**0.52**	(.03)	1.69	**0.51**	(.04)	1.67	**1.02**	(.03)	2.78	**1.48**	(.04)	4.40
norms favorable to drug use	**0.72**	(.03)	2.06	**0.98**	(.04)	2.66	**1.42**	(.03)	4.14	**1.76**	(.04)	5.79
laws favorable to drug use	**0.65**	(.02)	1.91	**0.85**	(.03)	2.33	**1.26**	(.03)	3.53	**1.59**	(.05)	4.92
perceived availability of drugs	**1.49**	(.04)	4.41	**1.80**	(.06)	6.03	**3.15**	(.06)	23.41	**3.37**	(.09)	28.96
perceived availability of guns	**0.37**	(.03)	1.44	**0.52**	(.03)	1.68	**0.74**	(.03)	2.09	**1.38**	(.04)	3.97
Family Risk												
parent attitude favorable to drugs	**2.34**	(.19)	10.40	**2.72**	(.21)	15.18	**3.03**	(.20)	20.64	**3.39**	(.21)	29.52
parent attitude favorable to ASB	**1.20**	(.07)	3.32	**1.35**	(.08)	3.85	**1.75**	(.08)	5.74	**2.17**	(.08)	8.75
School Risk												
academic failure	**0.77**	(.03)	2.17	**0.59**	(.03)	1.81	**1.24**	(.03)	3.44	**1.51**	(.05)	4.54
Peer / Individual Risk												
favorable attitudes toward ASB	**1.68**	(.05)	5.37	**1.88**	(.06)	6.52	**2.78**	(.07)	16.12	**3.63**	(.08)	37.83
favorable attitudes toward drugs	**3.49**	(.16)	32.92	**4.28**	(.21)	71.88	**5.44**	(.19)	230.67	**5.93**	(.21)	374.28
perceived risk of drug use	**0.97**	(.05)	2.62	**1.13**	(.05)	3.09	**1.83**	(.06)	6.25	**1.95**	(.08)	7.05
interaction with antisocial peers	**4.36**	(.21)	78.57	**4.25**	(.19)	69.97	**5.51**	(.20)	247.15	**6.18**	(.21)	482.51
friends drug use	**3.98**	(.17)	53.68	**4.72**	(.20)	111.61	**6.10**	(.20)	447.20	**6.45**	(.22)	635.24
rewards for antisocial involvement	**0.67**	(.03)	1.94	**0.88**	(.03)	2.40	**1.19**	(.04)	3.30	**1.60**	(.06)	4.93

Notes: Parameters indicated in bold are significat $p < .05$; OR is the Odds Ratio.

TABLE 6
Comparing Results from the 4-Class Exploratory and Replication Models

	Distributions		*4-Class–Exploratory*				*4-Class–Replication*			
	Exploratory	*Replication*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Proportion			50%	32%	14%	5%	62%	21%	10%	7%
Cigarettes										
Never	52.5%	64.0%	0.90	0.20	0.04	0.10	0.93	0.21	0.18	0.06
Lifetime	30.4%	23.5%	0.09	0.64	0.32	0.25	0.07	0.66	0.36	0.25
Current	17.1%	12.5%	0.01	0.16	0.64	0.65	0.00	0.14	0.46	0.69
Alcohol										
Never	40.0%	51.2%	0.73	0.10	0.01	0.06	0.75	0.23	0.01	0.05
Lifetime	30.4%	25.7%	0.22	0.54	0.13	0.15	0.20	0.56	0.06	0.14
Current	29.6%	23.1%	0.05	0.36	0.86	0.79	0.06	0.22	0.93	0.81
Marijuana										
Never	76.2%	83.1%	1.00	0.74	0.15	0.13	1.00	0.73	0.52	0.07
Lifetime	10.9%	7.5%	0.00	0.22	0.24	0.16	0.00	0.21	0.17	0.18
Current	12.9%	9.4%	0.00	0.05	0.60	0.71	0.00	0.06	0.31	0.76
Binge Drinking										
No	83.2%	87.0%	1.00	0.86	0.33	0.27	0.99	0.95	0.32	0.30
Yes	16.8%	13.0%	0.00	0.14	0.67	0.73	0.01	0.05	0.68	0.70
Suspended										
No	85.3%	88.0%	0.96	0.85	0.66	0.23	0.96	0.80	0.85	0.40
Yes	14.7%	12.0%	0.04	0.15	0.34	0.78	0.04	0.20	0.15	0.61

(*continued*)

TABLE 6
(*Continued*)

	Distributions		4-Class–Exploratory				4-Class–Replication			
	Exploratory	*Replication*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Carried gun										
No	93.0%	98.1%	0.98	0.95	0.61	0.27	0.98	0.94	0.95	0.69
Yes	7.0%	4.9%	0.02	0.05	0.09	0.73	0.02	0.06	0.06	0.32
Sold drugs										
No	93.5%	96.1%	1.00	0.99	0.78	0.22	1.00	0.99	0.98	0.46
Yes	6.5%	3.9%	0.00	0.01	0.22	0.78	0.00	0.01	0.02	0.54
Stole Vehicle										
No	95.8%	97.4%	1.00	0.99	0.92	0.40	1.00	0.98	0.98	0.69
Yes	4.2%	2.6%	0.00	0.02	0.08	0.60	0.00	0.02	0.02	0.31
Arrested										
No	92.5%	94.8%	1.00	0.95	0.81	0.26	1.00	0.94	0.99	0.50
Yes	7.5%	5.2%	0.01	0.05	0.19	0.74	0.00	0.06	0.04	0.50
Attacked someone										
No	83.7%	88.3%	0.97	0.84	0.60	0.12	0.97	0.83	0.80	0.34
Yes	16.3%	11.7%	0.04	0.17	0.40	0.88	0.03	0.17	0.20	0.66
High at school										
No	86.4%	90.5%	1.00	0.95	0.38	0.14	1.00	0.94	0.72	0.19
Yes	13.6%	9.5%	0.00	0.05	0.62	0.86	0.00	0.06	0.28	0.82
Gun to school										
No	97.9%	99.1%	1.00	1.00	0.99	0.57	1.00	1.00	1.00	0.87
Yes	2.1%	0.9%	0.00	0.00	0.01	0.43	0.00	0.00	0.00	0.13

Note: The solution reported here is for the random factor model.

prevalence of problem behaviors across data sets resulted in changes in the proportion of students in a given class across the two data sets and in the expected responses of students in the different class (see Tables 6 and 7). However, even though these values changed between the two samples, the interpretation of each class in the five-class model remained the same. For example, more than one third of the Problem Students in the replication sample reported carrying a gun to school in the last year and they reported similar levels as in the exploratory sample on nearly all of the other problem behaviors. The four-class model was not as well replicated in the 2000 data set (see Table 6). Here, the Drug Use class represented lower levels of substance use, especially marijuana use, and the Problem Behavior class also had lower levels of offending in the exploratory data.

Power to Detect Intervention Effects

A primary motivation for this article was to explore methodology for examining whether the CYDS can reduce the number of students in a community who experiment with substance use. By isolating Experimenters from those who engage in more serious levels of substance use and delinquency, we hope to test whether this universal intervention can reduce the likelihood that a student will experiment with substance use as distinct from reducing the probability of more serious problem behaviors. Analyses of the first three aims suggested that either a four- or a five-class model was appropriate for examining effects. Because these models differ largely in that the Experimenter class in the four-class model was split into an Experimenter and Alcohol Use class in the five-class model, we were concerned about a potential loss in power associated with having a smaller Experimenter class, so we first conducted power analyses to examine whether each model had sufficient power to detect an intervention effect on group membership. Power was assessed using Monte Carlo simulations, which involved generating 500 data sets from a population with known values. The proposed model was then fit with each of those data sets. Power was estimated as the percentage of times that the intervention effect was statistically significant ($p < .05$). The average standard error for the intervention effect and the sampling variance across all replications also were estimated. In theory, these two estimates should be very close to each other. One advantage of using Monte Carlo simulation was that it involved estimating the final model 500 times; any problems that are likely to occur when the final intervention data are used should appear in this process.

Data were generated for these Monte Carlo analyses using eight different true models. Because the CTC intervention might be expected to affect those students with low to moderate levels of problem behaviors, the treatment effect of interest was the regression of the latent class means for the Experimenter and Alcohol use

TABLE 7
Comparing Results from the 5-Class Exploratory and Replication Models

	5-Class–Exploratory					*5-Class–Replication*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
Proportion	51.4%	23.1%	11.1%	10.7%	3.7%	60.7%	19.5%	8.9%	8.5%	2.5%
Cigarettes										
Never	0.90	0.13	0.21	0.02	0.12	0.93	0.22	0.27	0.05	0.09
Lifetime	0.09	0.73	0.43	0.31	0.25	0.06	0.66	0.44	0.29	0.23
Current	0.01	0.13	0.36	0.67	0.63	0.00	0.12	0.29	0.66	0.68
Alcohol										
Never	0.71	0.14	0.00	0.01	0.07	0.75	0.26	0.01	0.04	0.05
Lifetime	0.23	0.69	0.07	0.14	0.16	0.20	0.60	0.05	0.15	0.13
Current	0.06	0.18	0.93	0.85	0.78	0.05	0.15	0.94	0.81	0.83
Marijuana										
Never	1.00	0.71	0.66	0.06	0.16	1.00	0.74	0.77	0.11	0.08
Lifetime	0.00	0.25	0.19	0.23	0.16	0.00	0.21	0.12	0.22	0.15
Current	0.00	0.05	0.16	0.71	0.68	0.00	0.05	0.10	0.67	0.77
Binge Drinking										
No	1.00	0.98	0.43	0.33	0.28	1.00	0.96	0.47	0.37	0.23
Yes	0.00	0.02	0.58	0.67	0.72	0.01	0.04	0.53	0.63	0.77
Suspended										
No	0.96	0.83	0.89	0.56	0.22	0.97	0.79	0.91	0.61	0.26
Yes	0.04	0.18	0.11	0.44	0.78	0.04	0.21	0.09	0.39	0.74

(*continued*)

TABLE 7
(*Continued*)

	5-Class–Exploratory					*5-Class–Replication*				
	1	*2*	*3*	*4*	*5*	*1*	*2*	*3*	*4*	*5*
Carried gun										
No	0.98	0.95	0.94	0.89	0.16	0.98	0.94	0.93	0.92	0.36
Yes	0.02	0.05	0.06	0.11	0.84	0.02	0.06	0.07	0.08	0.64
Sold drugs										
No	1.00	0.99	1.00	0.67	0.21	1.00	0.99	1.00	0.78	0.23
Yes	0.00	0.01	0.00	0.33	0.79	0.00	0.01	0.00	0.22	0.77
Stole Vehicle										
No	1.00	0.98	0.98	0.89	0.33	1.00	0.98	0.99	0.91	0.42
Yes	0.00	0.02	0.02	0.11	0.67	0.00	0.02	0.01	0.09	0.58
Arrested										
No	1.00	0.94	0.97	0.72	0.25	1.00	0.94	0.99	0.77	0.28
Yes	0.01	0.06	0.03	0.28	0.76	0.00	0.06	0.01	0.23	0.72
Attacked someone										
No	0.93	0.83	0.81	0.52	0.10	0.97	0.83	0.88	0.61	0.11
Yes	0.04	0.17	0.19	0.49	0.90	0.03	0.17	0.15	0.39	0.89
High at school										
No	1.00	0.95	0.87	0.25	0.15	1.00	0.96	0.90	0.34	0.12
Yes	0.00	0.05	0.13	0.75	0.85	0.00	0.05	0.10	0.66	0.88
Gun to school										
No	1.00	1.00	1.00	0.99	0.47	1.00	1.00	1.00	1.00	0.66
Yes	0.00	0.00	0.00	0.01	0.53	0.00	0.00	0.00	0.00	0.34

Note: The solution reported here is for the random factor model.

classes on the treatment status of the community. Lower values for the mean of the Experimenter class in treatment communities indicate that students in those communities are more likely to be in the Abstainer class than the Experimenter class. Power of the intervention to move students from the Experimenter to the Abstainer class was evaluated for the four-class model. For the five-class model, power was evaluated to move students from the Experimenter to the Abstainer class (holding the other classes constant), from the Alcohol Use to the Abstainer class (holding the other classes constant), and from both the Experimenter and Alcohol Use classes to the Abstainer class (with the same effect size for each class). For each of these four conditions, power was estimated for two effect sizes corresponding to the ability of the intervention to move 25% and 35% of the students from each class to the Abstainer class. These values were chosen to represent small and medium effects; other community level prevention studies have found reductions in substance use of between 20% and 40% (Perry, Williams, Komro, & Veblen-Mortenson, 2000; Perry et al., 1996). Further, the CYDS study was powered to detect intervention effects of this size (Murray, Van Horn, Hawkins, & Arthur, 2006), so it makes sense to evaluate whether these models have comparable power to multilevel general linear models.

Data generation used the results from the 1998 data set as the starting point for parameter estimates but included only 24 communities as in the CYDS study. Data were generated separately for the 12 communities in the intervention and for the 12 communities in the control condition. Each of the 24 communities for which data were generated had a different number of students, matching the number of students for whom data were collected in the first wave of the CYDS. Item thresholds and the proportion of students in each class in the control condition were from the 1998 results. The data generation model for the treatment condition was identical except that the class means corresponding to a 25% or 35% movement of students from the affected class(es) to the Abstainer class were used instead of the observed class means. It should be noted here that a movement of students from one class to the reference class resulted in a change in the means for each class because the means are the odds of being in one class versus the reference class. The movement of students between these classes changes the number of students in the reference category and thus changes all of the odds. After 500 data sets for both intervention and control communities were generated for each of the eight true models, communities from the two conditions were combined and the final model was run on each of the 500 merged data sets. Data was generated for the five-class model such that the intervention reduced both the Experimenter and Alcohol Use classes, and two models were run. In one model, the effects were estimated separately and in the second, intervention effects were constrained to be the same for the two classes. The rationale for this constraint was that, theoretically, it may be expected that the intervention will affect both of these classes similarly. We wanted to evaluate whether imposing this constraint on the model would result in an increase in power. The cost of imposing this constraint is that, if the intervention affects one group more than the other group, the effect may not be detected. Running the model both ways allowed for an evaluation of the possible gain made by this trade-off. One final issue is that with 24 communities these models have more fixed effects (item thresholds are all estimated as fixed) than clusters. Although theoretically this should not be a problem, it does result in the first order derivative product matrix being nonpositive definite, which causes an error message in Mplus. Additional work is needed to evaluate the effects of having more fixed effects than clusters.

Results from the power analyses in Table 7 indicated that CYDS has fair to good power to detect an intervention effect of 35%. In the four-class model, power to detect a movement of 35% of the experimenters to the Abstainer class is .91. In the five-class model, power is about .75 to detect a 35% reduction in either the Experimenter or Alcohol Use classes. However, if both classes are positively affected by the intervention, power is increased, and if both of the intervention effects are constrained to be the same, power to detect both of the effects together is .93. Thus, so long as the intervention effect is the same for the two groups, the constrained model results in an increase in power and a reduction in the degrees of freedom used to test the intervention effects. As expected, power is lower when the effect size of the intervention is a more modest 25% movement in students from a given class to the Abstainer class. Power for these models ranges from .42 to .67.

Effects of the CYDS Intervention

The final aim of this article is to demonstrate the use of multilevel mixture models to examine effects of the CYDS intervention on class membership. Analyses were performed using data collected in 2004 after the 1st year of the study. The 1st year involved community planning rather than the implementation of interventions, so no differences were expected between treatment and control communities in the proportion of students in each class. These analyses test whether there are differences between the treatment and control communities in the proportion of students identified as experimenters and/or alcohol users as compared with those identified as abstaining from substance use. As depicted in Figure 1, the effect of the intervention is on class membership, and the random-factor model from Figure 2 was used to capture between-community differences in the proportion of students in each class.

Because of the large number of thresholds (estimated as fixed effects at the community level) and the small number of communities typically involved in group randomized trials, we suggest a specific approach to building multilevel mixture models. We first estimated fixed-effects models to identify class thresholds in the four- and five-class models. Next, we estimated random-effects models in which the item thresholds were constrained based on the results of the fixed-effects estimates. The final step was to estimate the thresholds and random effects simultaneously. The point of this building process is to establish that both the fixed- and random-effects portions of the model are stable. Estimates of the thresholds from the first model should be very similar to estimates from the third model and estimates of the random effects from the second model should be similar to those obtained in the third model.

The results from our fixed-effects model were substantively the same as those from the 1998 sample. The four-class model found Abstainer (58%), Experimenter (26%), Drug Use (12%), and Problem Student (4%) classes, and the five-class model identified Abstainer (59%), Experimenter (20%), Alcohol Use (8%), Drug Use (9%), and Problem Student (3%) classes. The second step involved estimating a multilevel mixture model in which estimates of thresholds from the fixed-effects model were used to fix the thresholds. In the final step, the within-level thresholds and between-level random factor model were both estimated and a treatment effect on the latent class means was included. To establish the stability of the model, the thresholds from the first model were compared with those from the third model and were found to be very similar, never more than one half of a standard error apart. Differences between the random effects in the second step and third step were also evaluated and no differences in parameter estimates were found. However, standard errors for the random effects and intervention effects were up to 20% larger in the analyses with all parameters estimated. Further work should be conducted to evaluate which standard errors are more accurate.

To assess differences between intervention and control communities, the latent class means for each community were regressed on treatment status. Because the class means capture the proportion of students in a community in a given class relative to the reference class (Abstainers), negative treatment effects indicate that there are fewer students in that class relative to Abstainers in the treatment than in the control communities. Because the power analyses showed a clear benefit of constraining the Experimenter and Alcohol Use classes to have the same effect and, given that we do not expect differences in effects between these classes, those constraints were imposed in these analyses. Results for both the four- and five-class models are depicted in Table 8. As expected, none of the intervention effects are significantly different from zero, serving to confirm that these communities are quite similar in the proportion of students in each class in 2004, before the intervention focused on youth behaviors began in the CTC communities. For purposes of illustration we interpret the results although they are not significantly different from zero. From the four-class model, students in communities in the treatment condition have an odds of being in the Experimenter versus Abstainer class of .73 times that of students

TABLE 8
Power to Detect Treatment Effects for Different Models

	25% Reduction			*35% Reduction*		
Model/class affected	*Beta*	*SE*	*Power*	*Beta*	*SE*	*Power*
4-class						
experimenter	−0.435	0.182	0.480	−0.627	0.183	0.906
5-class						
alcohol use	−0.342	0.189	0.428	−0.514	0.200	0.754
5-class						
experimenter	−0.407	0.212	0.424	−0.588	0.215	0.742
5-class–both e&a						
alcohol	−0.443	0.178	0.666	−0.643	0.181	0.920
experimenter	−0.447	0.209	0.518	−0.644	0.212	0.816
5-class—both e&a						
e&a (constrained)	−0.444	0.174	0.670	−0.642	0.176	0.934

Note: The SE is the average SE across all simulations, e&a is experimenter and alcohol use.

TABLE 9
Effects of the CYDS Intervention on Class Membership in 2004

	4-Class			*5-Class*		
	Beta	*SE*	*OR*	*Beta*	*SE*	*OR*
Experimenter	−0.29	0.18	0.75	−0.23	0.21	0.79
Alcohol Use	n/a	n/a	n/a	−0.23	0.21	0.79
Drug Use	0.14	0.25	1.15	0.25	0.33	1.28
Problem Students	−0.18	0.26	0.84	−0.21	0.19	0.81

Note: Analyses include 5111 8th grade students from 24 communities.

in the control condition. From the five-class model, students in communities in the treatment condition have an odds of .77 times that of students in the control condition of being in either the Alcohol Use or Experimenter classes.

DISCUSSION

This article evaluated the use of multilevel mixture models to detect the effects of interventions in group randomized trials. This was accomplished using the Community Youth Development Study as an example of how to estimate the number of latent classes, specify between-cluster differences, evaluate validity of the classes obtained, replicate those classes on an independent data set, use the resulting data to estimate power to detect intervention effects for each class, and assess effects of the intervention on the proportion of students in each latent class after the planning phase of the study. In the discussion, we review the results of these analyses in the context of the CYDS, comment on their application to other group randomized trials, and suggest future areas of research.

Conclusions for CYDS

The model proposed here for assessing the effects of CYDS answers the research question, “Does the Communities That Care intervention reduce the probability that students will engage in experimental use of illicit drugs and/or engage in high levels of alcohol use?” The research question itself was partially determined by the results of the initial analyses, which examined the number and nature of the classes capturing differences among students in substance use and delinquency. The analyses first examined different solutions for the latent class portion of the model and found that fit criteria and substantive interpretation of the classes suggested the use of a four-class or five-class solution. Further work examined the validity and replicability of these two alternatives. The five latent classes include Abstainers, Experimenters, Alcohol Users, Drug Users, and Problem Students. The four-class model differed in that Alcohol Users and Experimenters were combined into one class. The model also estimated random effects by evaluating differences between schools in the proportion of students in each class. These results show clearly that the probability that a student would be in a given class differs between schools and that those differences vary across latent classes. Schools differ the most with respect to the prevalence of Problem Students; in the five-class solution the odds of a student being in this class versus the Abstainer class in schools at the 10th percentile was .02 whereas in schools at the 90th percentile the odds were .13. Results of these analyses further suggested that the random-factor model adequately captured between-school differences.

These initial analyses used two data sets collected prior to the CYDS and contained a larger number of students/communities/clusters. The availability of these data was a major advantage as they included far more clusters than were available in the CYDS group randomized trial. Thus, the latent classes and estimates of between-cluster variance should be more stable and more replicable. These data sets provided a strong basis for the estimation of power to detect effects in CYDS. Power analyses demonstrated first that with the appropriate constraints placed on the model, multilevel mixture analyses could be used to assess intervention effects in CYDS. The final Monte Carlo simulations also found that the proposed models had 93% power to detect a 35% reduction of Experimenters and Alcohol users and 67% power to detect a reduction of 25%. In this case, use of the mixture model did not seem to come at the cost of a loss in power compared with the use of other models (Murray et al., 2006). Further, results of the power analyses demonstrated that constraining the effects of the intervention on two classes to be the same helped preserve power. This is an important finding because researchers may be inclined to select a model with fewer classes in hopes of increasing power to find effects. The ability to constrain effects to be the same across classes allows the flexibility to select the optimal number of classes and then increase power by imposing theoretically based constraints.

Finally, this article demonstrated that it is possible to estimate these models with group randomized trials by regressing class membership on community treatment status after the 1st year of the CYDS intervention. We note that this model presumes that the treatment does not in fact affect the composition of the latent classes. If the latent classes themselves differ between treatment and control communities, it is not possible to assess the effects of the intervention on class membership. Although not reported, we did test for differences in class composition between treatment and control communities in our analyses by running the latent class analyses separately for the treatment and control arms, finding that the same classes emerge from communities in both conditions. The findings of no difference between treatment and control communities in the proportion of students in each class provided evidence for the comparability of the two conditions with respect to the profiles of problem behaviors before interventions focused on changing young people's behaviors began in CYDS and illustrated methodology for assessing affects later in the study.

Implications for the Use of Mixture Models to Evaluate Group Randomized Trials

This work has broader implications for the assessment of GRTs. First, it indicates that multilevel mixture models are a strong alternative for assessing intervention effects on different groups or profiles of students. The models used in this article are multilevel latent class analyses, a subset of finite mixture models. Although the use of latent class analysis is well established in the social sciences, the expansion of these models to multilevel data is new and has implications that should be more fully explored. There are clear advantages of using latent classes to identify groups of students who may be differentially affected by interventions. Latent categorical variables are able to correct for measurement error and explicitly model class membership as opposed to a procedure in which respondents are classified based on some cut point of the outcome variables.

This study also suggests a specific analytic process for using multilevel mixture models to examine differential effects of preventive interventions. The first step involves assessing the fixed effects, determining the number of classes to be used, and interpreting those classes. The second step centers on assessing random effects and ascertaining the appropriate model specification to allow between-cluster variation. In these analyses, the inclusion of random effects did not change the fixed-effects portion of the model, but that may not always be the case. Thus, the fixed effects should be reexamined with the final random-effects specification. Third, we recommend assessing the validity of the results obtained and replicating the results on an independent sample. Finally, the Monte Carlo simulations reported provide an effective way to assess power in future evaluations. The example used here asked whether intervention affected the probability of engaging in experimental substance use or regular alcohol use. Other future analyses could examine interactions between the latent class variable and an independent outcome, which would assess whether the intervention effects differ across classes.

Finally, we note that although this study addresses the use of multilevel mixtures to test group randomized trials, the results can be applied to any situation in which clustering occurs, as the intervention status variable is simply a cluster-level variable. For example, those studying effects of a school- or community-level variable can use these models to examine individual by context interactions.

Future Directions

Further work is needed to expand the use of multilevel mixtures. One area for future efforts involves developing techniques to analyze longitudinal data at the community level. In the CYDS, for example, cross-sectional data are being collected from each community every 2 years. Ideally, we would like to assess whether the intervention affects a change in the proportion of students in a class rather than assessing differences in class membership at posttest. Additional research should examine how this might be accomplished. Also, it would be useful to assess the effects of model misspecification on the estimates of intervention effects. For example, if the four-class rather than five-class model were used, how would the intervention effects differ? If a fixed-effects rather than random-effects model were used, would the standard errors be biased as expected?

This article demonstrates how these models can be used in a multilevel context to better understand the effects of interventions targeting students. We believe that these modeling techniques provide a potentially powerful tool for assessing intervention effects. We are only beginning to investigate the many ways that these models may be useful for understanding human behavior.

ACKNOWLEDGMENTS

This research was supported by Research Grant 5R01DA015183-03 from the National Institute on Drug Abuse awarded to J. David Hawkins. We are grateful to the co-funding organizations: the National Cancer Institute, the Center for Substance Abuse Prevention, the National Institute on Mental Health, and the National Institute on Child Health and Development. We are also grateful for the support and feedback from our colleagues in the Prevention Science and Methodology Group, supported by the National Institute of Mental Health and the National Institute on Drug Abuse through Grant 5R01MH40859. We also appreciate the assistance of Katherine Masyn and Charlie Fleming.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716–723.

Allen, J. P., & Philliber, S. (2001). Who benefits most from a broadly targeted prevention program? Differential efficacy across populations in the Teen Outreach Program. *Journal of Community Psychology, 29*(6), 637–655.

Arthur, M. W., Hawkins, J. D., Pollard, J. A., Catalano, R. F., & Baglioni, A. J. (2002). Measuring risk and protective factors for substance use, delinquency, and other adolescent problem behaviors—The Communities That Care Youth Survey. *Evaluation Review, 26*, 575–601.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*, 3–29.

Botvin, G., Mihalic, S., & Grotpeter, J. K. (Eds.). (1998). *Life Skills Training* (Vol. 5). Boulder: Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado.

Clark, L. A., & Watson, D. (1995). Construct validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity on psychological test. *Psychological Bulletin, 52*, 281–302.

Diehr, P., Martin, D. C., Koepsell, T., & Cheadle, A. (1995). Breaking the matches in a paired t-test for community interventions when the number of pairs is small. *Statistics in Medicine, 14*, 1491–1504.

Eisen, M., Zellman, G. L., Massett, H. A., & Murray, D. M. (2002). Evaluating the Lion's-Quest "Skills for Adolescence" drug education program: First-year behavioral outcomes. *Addictive Behaviors, 27*, 619–632.

Ellickson, P. L., & Bell, R. M. (1990). Drug prevention in junior high: A multi-site longitudinal test. *Science, 247*, 1299–1305.

Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*(1), 47–53.

Feng, Z., Diehr, P., Paterson, A., & McLerran, D. (2001). Selected statistical issues in group randomized trials. *Annual Review of Public Health, 22,* 167–187.

Feng, Z., Diehr, P., Yasui, Y., Evans, B., Beresford, S., & Koepsell, T. D. (1999). Explaining community-level variance in group randomized trials. *Statistics in Medicine, 18*, 539–556.

Frey, K. S., Hirschstein, M. K., Snell, J. L., Edstrom, L. V. S., MacKenzie, E. P., & Broderick, C. (2005). Reducing playground bullying and supportive beliefs: An experimental trial of the Steps to Respect program. *Developmental Psychology, 41*(3), 479–491.

Glaser, R. R., Van Horn, M. L., Arthur, M. W., Hawkins, J. D., & Catalano, R. F. (2005). Measurement properties of the Communities That Care Youth Survey across demographic groups. *Journal of Quantitative Criminology, 21*, 73–102.

Hawkins, J. D. (2004). *Communities That Care prevention strategies guide*. South Deerfield, MA: Channing Bete.

Hawkins, J. D., & Catalano, R. F. (1992). *Communities That Care: Action for drug abuse prevention*. San Francisco: Jossey-Bass.

Hawkins, J. D., Catalano, R. F., & Arthur, M. W. (2002). Promoting science-based prevention in communities. *Addictive Behaviors, 27*, 951–976.

Hawkins, J. D., Catalano, R. F., Arthur, M. W., Egan, E., Brown, E. C., Abbott, R. D., et al. (in press). Testing Communities That Care: Rationale and design of the Community Youth Development Study. *Prevention Science.*

Hawkins, J. D., Van Horn, M. L., & Arthur, M. W. (2004). Community variation in risk and protective factors and substance use outcomes. *Prevention Science, 5*(4), 213–219.

Hedeker, D. (2003). A mixed effects multinomial logistic regression model. *Statistics in Medicine, 22*, 1433–1446.

Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11*, 36–53.

Kellam, S. G., Ling, X., Merisca, R., Brown, C. H., & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology, 10,* 165–185.

Lochman, J. E., Curry, J. F., Dane, H., & Ellis, M. (2001). The Anger Coping Program: An empirically-supported treatment for aggressive children. *Residential Treatment for Children and Youth, 18*(3), 63–73.

Loeber, R., & Farrington, D. P. (1998). *Serious and violent juvenile offenders: Risk factors and successful interventions*. Thousand Oaks, CA: Sage.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Mihalic, S., Fagan, A., Irwin, K., Ballard, D., & Elliott, D. (2004). *Blueprints for violence prevention (NCJ 204274).* Washington, DC: Office of Juvenile and Delinquency Prevention.

Moffitt, T. E. (1993). Adolescent-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674–701.

Murray, D. M. (1998). *Design and analysis of group-randomized trials* (Vol. 27). New York: Oxford University Press.

Murray, D. M., Clark, M. H., & Wagenaar, A. C. (2000). Intraclass correlations from a community-based alcohol prevention study: The effect of repeat observations on the same communities. *Journal-of-Studies-on-Alcohol, 61*(6), 881–890.

Murray, D. M., Feldman, H. A., & McGovern, P. G. (2000). Components of variance in a group-randomized trial analyzed via a random-coefficients model: The Rapid Early Action for Coronary Treatment (REACT) trial. *Statistical Methods in Medical Research, 9*, 117–133.

Murray, D. M., Van Horn, M. L., Hawkins, J. D., & Arthur, M. W. (2006). Analysis strategies for a community trial to reduce adolescent ATOD use: A comparison of random coefficient and ANOVA/ANCOVA models. *Contemporary Clinical Trials, 27*, 188–206.

Muthén, B. O., Brown, C. H., Masyn, K., Jo, B., Khoo, S., Yang, C., et al. (2002). General growth mixture modeling for randomized prevention trials. *Biostatistics, 3*, 459–475.

Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55*, 463–469.

Muthén, L. K., & Muthén, B. O. (2004). Mplus (Version 3). Los Angeles: Author.

Muthén, L. K., & Muthén, B. O. (2006). Mplus (Version 4). Los Angeles: Author.

Nylund, K. L., Asparauhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569.

Perry, C. L., Williams, C. L., Komro, K. A., & Veblen-Mortenson, S. (2000). Project Northland high school interventions: Community action to reduce adolescent alcohol use. *Health Education and Behavior, 27*(1), 29–49.

Perry, C., Williams, C. L., Komro, K. A., Veblen-Mortenson, S., Stigler, M. H., Munson, K. A., et al. (2002). Project Northland: Long-term outcomes of community action to reduce adolescent alcohol use. *Health Education Research, 17*(1), 117–132.

Perry, C. L., Williams, C. L., Veblen-Mortenson, S., Toomey, T. L., Komro, K. A., Anstine, P. S., et al. (1996). Project Northland: Outcomes of a community wide alcohol use prevention program during early adolescence. *American Journal of Public Health, 86*, 956–965.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods, 5,* 199–213.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*, 333–343.

Sherman, L. W., Gottfredson, D. C., Mackenzie, D., Eck, J., Reuter, P., & Bushway, S. (Eds.). (1997). *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress*. Washington, DC: U.S. Department of Justice, Office of Justice Programs.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Stoolmiller, M., Eddy, J. M., & Reid, J. B. (2000). Detecting and describing preventative intervention effects in a universal, school-based randomized control trial targeting delinquent and violent behavior. *Journal of Consulting and Clinical Psychology, 68*, 296–306.

U.S. Department of Health and Human Services. (2001). *Youth violence: A report of the surgeon general*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Mental Health Services, National Institutes of Health, National Institute of Mental Health.

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology, 33*, 213–239.