A mixture IRT analysis of risky youth behavior

W. Holmes Finch* and Eric E. Pierson

Department of Educational Psychology, Ball State University, Muncie, IN, USA

The study reported in this manuscript used a mixture item response model with data from the Youth Risk Behavior Survey 2009 (N = 16,410) to identify subtypes of adolescents at-risk for engaging in unhealthy behaviors, and to find individual survey items that were most effective at identifying such students within each subtype. The goal of the manuscript is twofold: (1) To demonstrate the utility of the mixture item response theory model for identifying subgroups in the population and for highlighting the use of group specific item response parameters and (2) To identify typologies of adolescents based on their propensity for engaging in sexually and substance use risky behaviors. Results indicate that four classes of youth exist in the population, with differences in risky sexual behaviors and substance use. The first group had a greater propensity to engage in risky sexual behavior, while group 2 was more likely to smoke tobacco and drink alcohol. Group 3 was the most likely to use other substances, such as marijuana, methamphetamine, and other mind altering drugs, and group 4 had the lowest propensity for engaging in any of the sexual or substance use behaviors included in the survey. Finally, individual items were identified for each group that can be most effective at identifying individuals at greatest risk. Further proposed directions of research and the contribution of this analysis to the existing literature are discussed.

Keywords: mixture item response theory, latent class analysis, item response theory, mixture modeling, at-risk youth, youth risk behavior survey

INTRODUCTION

For many individuals, adolescence is a period of time marked by exploration, self-discovery, and risk taking. For most, these activities will eventually lead to positive outcomes with nothing more costly than a few exciting memories and tales of misadventure (Berk, 2008). For some individuals, however, the cost of experimentation and risk taking behavior will lead to problem drug and alcohol use, sexual victimization, peer victimization, serious health injury, unexpected pregnancy, additional mental health problems, or death (Harley et al., 2010). These outcomes may result in significant costs to the individual in the form of mental and physical injury, reduced academic progress, and a reduction in the quality of life. In addition, family and friends of the individual may experience a loss of productivity in the workplace, conflict with significant others, and financial stress in helping to support adolescents who have been involved in these problematic activities (Hamrin et al., 2010).

When parents become aware of certain types of risk taking behavior they may turn to mental health professionals to determine whether or not the behavior is dangerous or merely an act of youthful exploration. One way of considering this question is by assessing the likelihood that an adolescent who has engaged in a particular type of risky behavior will escalate into more such activities. For example, is an individual who experiments with smoking tobacco at greater risk for using other, more serious substances such as heroin? Because adolescence typically involves a wide range of exploratory activities, there is a great incentive to develop an efficient and accurate means of identifying harmful behaviors and to intervene early with youth who are exhibiting these behaviors. One common tool for gathering information about risky behaviors is through the use of surveys and questionnaires asking about specific behaviors.

A prime example of such an instrument is the youth risk behavior survey (YRBS), which is conducted by the U.S. Centers for Disease Control and Prevention (CDC). This instrument, which will be discussed in greater detail below, asks adolescent respondents to report on their behaviors in a variety of areas, including sexual, substance use, violence, bullying, suicidal thoughts and actions, and healthy eating and exercise. The use of such a questionnaire can lead to the identification of a complex series of relationships among behaviors that varies across subgroups in the population (Bekman et al., 2010) and across ethnicities (Shih et al., 2010). Such complex relationships among behaviors may not be easy to discern, particularly for a large sample of people. In addition, individual items or behaviors may exhibit differential effectiveness across population subgroups in identifying individuals at-risk for escalating their dangerous behaviors. In an attempt to address this complexity, the current study was conducted to better understand the nature of subgroups in the population of adolescents, based on their risky behaviors, and to identify item response patterns unique to these groups that would be helpful in singling out those at greatest risk. The analysis used to achieve these goals is the mixture item response theory (mixIRT) model, which will be described in some detail next. Prior to this description, we first review some recent research on the identification of at-risk adolescents.

Prior research in the identification of adolescents prone to risky behaviors has typically relied on the use of existing instruments or scales that assumed to be equally valid for all members of the population. For example, Prado et al. (2010) used a structural equation model in conjunction with a variety of measures of adolescent–parent relationships, substance use, and school performance to develop a model for predicting early sexual activity

among Hispanic youth. Coley et al. (2009) utilized a latent growth model to examine the link between parent involvement and early onset of sexual activity in adolescents, and found that the more knowledge parents had regarding their children's activities outside of school, the later the adolescents first engaged in sex. Prediction of substance abuse in adolescents has been studied in a number of cases, including recent work by Lee et al. (2010), who used latent class regression analysis to predict which types of individuals were most at-risk to abuse a variety of drugs and alcohol. Results of this study identified several risk factors including being male, dropping out of high school, and being diagnosed with conduct disorder or oppositional defiant disorder were all predictive of substance abuse in adolescence and early adulthood. Other researchers (Reininger et al., 2005) used multiple linear regression analysis to predict an aggregate risk score comprised of smoking, drinking and number of sex partners, using a variety of assets available to adolescents. Their results showed that adolescents with greater peer encouragement to avoid risky behaviors and more school support had the lowest risk scores. Finally, Vierhaus and Lohaus (2008) found that parent and child reports of psychopathologic symptoms in sixth grade were predictive of adolescents engaging in risky sexual and substance use behavior in high school.

As mentioned earlier, much of this previous research has focused on the use of total scores on measures of risk and predictors of risk. The current study will make use of item level information from the YRBS to identify groups of adolescents who might be at-risk for self injurious substance use and sexual behaviors. In addition, the mixIRT model used in this research will provide information regarding which behaviors might be the most useful for identifying individuals prone to be at-risk within each of these classes.

MIXTURE ITEM RESPONSE THEORY MODEL

The mixIRT model combines two powerful statistical tools that have been used extensively, but separately, in the Psychological sciences, latent class analysis (LCA) and item response theory (IRT). The use of the mixIRT model in a variety of contexts has been described in detail by a number of authors (Yamamoto, 1987; Mislevy and Verhelst, 1990; Rost, 1990; von Davier and Rost, 1995; von Davier and Yamamoto, 2004; Cohen and Bolt, 2005). It has been recommended for identifying subsets of the population (latent classes) that are characterized by different item response models for a particular measure or instrument (Li et al., 2009). In this context, psychometricians have used it to detect and characterize differential item functioning (DIF), for example (e.g., Bolt et al., 2001, 2002; De Ayala et al., 2002; Cohen and Bolt, 2005).

Much (though not all) of this earlier research has used the mixIRT model in the realm of educational measurement. In contrast, the focus of the current study was on using a mixIRT model to characterize population subgroups of adolescents with respect to their propensity to engage in risky sexual and substance use behaviors, and to find individual behaviors that might be particularly useful for identifying individuals at greater risk for escalating into more dangerous activities, within each of the subpopulations. Some prior research has been conducted using mixture models and LCA with personality inventories and other affective measures (e.g., Eid, 1997; Frick et al., 1997; Fossati et al., 2001; Smit et al., 2003) and to identify individuals at-risk for exhibiting problem behaviors. As an

example of the latter, Uebersax (1997) used latent mixture models and probit latent class models to identify classes of adolescent males based on problem behaviors such as fighting, destruction of property, and drinking and driving. This study found four classes, those who are prone to engage in a wide range of problem behaviors, those unlikely to engage in any such behaviors, those likely to fight, damage property and have dealings with the police, and those likely to have problems at school. Sorenson et al. (1997) used a variety of latent variable models, including a mixture Rasch approach with six items taken from Hirschi's (1969) scale of delinquent behavior used in the Seattle Youth Study. The behaviors all involved crimes involving fighting or property damage/theft. Their goal was to define a latent variable of adolescent delinquent behavior, which they were able to do though it was found to differ between males and females. In the realm of identifying groups based on anti-social behavior, van der Heijden et al. (1997) used LCA 24 dichotomous items given to 2918 individuals between the ages of 12 and 24. These items asked respondents whether they had engaged in a variety of behaviors involving theft, arson, assault, and vandalism. They identified five latent classes, which could be characterized as not engaging in anti-social behavior, younger males prone to violence and vandalism, younger females prone to theft and vandalism, older males prone to violence and vandalism, and finally those of both genders prone to the entire range of anti-social behaviors.

The two parameter logistic mixIRT model (mix2PL) for dichotomous data, which was used in this study, is an extension of the mixture Rasch model described by Rost (1990) and is expressed as

$$P(U = 1 \mid g, \theta_g) = \frac{\exp\{a_{ig}(\theta_g - b_{ig})\}}{1 + \exp\{a_{ig}(\theta_g - b_{ig})\}}. \qquad (1)$$

Here latent class membership ($g = 1, 2, \ldots, G$), within class item difficulty (b_{ig}), within class item discrimination (a_{ig}), and the within class level on the latent trait being measured (θ_g) are all model parameters to be estimated. In addition, each survey respondent is placed in a latent class, and the proportions of individuals in each class (π_g), are also estimated, under the constraint that $\sum_{g=1}^{G} \pi_g = 1$. The item difficulty and discrimination parameters carry the same meaning in the mixIRT context as they do in the more general IRT framework. Thus, for example, in the context of cognitive assessments and achievement testing difficulty provides information regarding the likelihood that an individual will answer an item correctly, while discrimination indicates how well the item differentiates between individuals with different levels of the construct being measured (De Ayala, 2009). Item difficulty is scaled so that large positive numbers indicate that the item is relatively more difficult while large negative numbers indicate that the item is relatively easy. Item discrimination should generally be positive, with larger values indicating that the item is better able to distinguish individuals with different levels of the latent construct being measured. In the current study the included items ask students to report whether they engaged in a variety of sexually risky behaviors and substance use. Therefore, the latent trait being measured was the propensity for engaging in such behaviors. Thus, higher values of the latent trait would indicate that an individual is relatively more at-risk, and similarly items (behaviors) with higher difficulty values were less likely to be endorsed, while those with

higher discrimination values were better able to distinguish among individuals with different levels of being at-risk. The mix2PL model was selected rather than a mix1PL or mix3PL because of the goals of this study and the nature of the data we are using. Specifically, a primary goal of this research was to identify items that performed differently for different at-risk classes within the broader population. Because educators and mental health professionals use this scale to identify adolescents at-risk for problem behaviors, both item location and item discrimination are important. The latter parameter is particularly informative in this regard because it provides information regarding which items are best (and worst) at identifying potentially at-risk youth. For this reason, the mix1PL model did not seem appropriate, as it holds discrimination constant across items. Furthermore, the nature of the questions being asked on the YRBS, which focus on respondent's participation in specific behaviors, would seem to make chance endorsement of an item highly unlikely and thus make the mix3PL model a less than optimal analysis choice as well.

When there are class differences in the item difficulty and discrimination parameter values, it is possible to conclude that members of different classes perceive the item differently. For example, if b has a higher value for latent class 1 versus class 2, we know that the item is more difficult for class 1, which in turn may provide insights into the types of individuals who tend to be in that class. Similarly, if class 2 has a higher discrimination parameter value on a given item than does class 1, that item is better able to differentiate among individuals with different levels of the latent trait for class 2 than for class 1. This approach to using mixIRT models has been particularly evident in the identification and characterization of DIF for achievement tests (Cohen and Bolt, 2005), though it has also been used to identify different usage patterns of the "Not Sure" category in personality inventories (Maij-de Meij et al., 2008) and to identify individuals engaging in impression management in organizational surveys (Eid and Zickar, 2007).

Parameter estimation for mixIRT models can be carried out either using maximum likelihood (MLE) or through the use of Markov-chain Monte Carlo (MCMC) methods in the Bayesian context (von Davier and Rost, 2007). Each of these approaches has its advantages and disadvantages in practice. For example, MCMC has proven useful in the estimation of complex mixIRT models because it does not require the integration of the likelihood function, which can be extremely difficult when many parameters must be estimated (Junker, 1999). On the other hand, the MCMC approach is often very time consuming to implement (sometimes taking 10 days or more to fit a single model), and may encounter difficulties in converging to solutions for individual parameters. This issue of time is non-trivial when dealing with mixIRT models, as several different latent class solutions must typically be fit and then compared in order to determine which is optimal for the data at hand (Li et al., 2009). MLE does not typically require the time that MCMC does, and has also been used successfully in estimating mixIRT models (von Davier and Rost, 2007). However, because MLE can mistakenly converge on localized rather than general MLE solutions, leading to suboptimal model parameter estimates, it is important to use multiple random starting values, such as the 10 random starts used in the current study (Rost, 1991).

GOALS OF THE CURRENT STUDY

The current study has two major goals. First of all, it seeks to explore and demonstrate how mixIRT models can be used to identify subgroups within a population based on a behavior inventory, and how different behaviors may be more useful for identifying at-risk youth within specific subgroups. Secondly, the mix2PL model was applied to items from the YRBS specifically to identify subtypes of individuals at varying degrees of risk for engaging in sexual and substance use behaviors, and to determine which of the items (behaviors) might be most effective at identifying such at-risk individuals within each subgroup.

MATERIALS AND METHODS
MEASURE

The YRBS is a biennial survey conducted by the United States CDC designed for high school students. It consists of self-reports of past and current thoughts, emotions, behaviors, and exposure to health relevant curriculum within schools, and provides a snapshot of the general health and risky behavior of America's youth (Centers for Disease Control, 2009). The YRBS dataset is publicly available with individual identifying information, as well as school, state, and regional identifiers removed. This anonymity is crucial in order for respondents to feel comfortable responding to items regarding risky sexual behaviors, substance use, suicidal thoughts, and violent activities that they may otherwise be reluctant to answer honestly and openly without such guarantees. While typically validity is increased by the evaluation of behaviors, thoughts, and attitudes from multiple perspectives, the nature of the questions asked on the YRBS would likely lose sensitivity to the frequency, duration, and acknowledgment of the behaviors (Sattler and Hoge, 2008). This is particularly true of behaviors that may invite unwanted attention on the part of peers and adults (McConaughy and Ritter, 2008; Sattler and Hoge, 2008).

The YRBS contains items that are used to create more than 200 variables in the final dataset. As stated previously, the goal of the present study was to better understand the latent structure of common core psychological and emotional characteristics present in the adolescent population, with respect to particular sexually risky and substance use behaviors. To this end, 14 items were selected (see **Table 1**) from the YRBS that inquire about such behaviors in a dichotomous format; e.g., respondents were coded as either engaging in the behavior or not. These dichotomous items were created by the CDC from the original set of items, which asked respondents to indicate the first onset, frequency or duration of risky behaviors. As an example, in this study the dichotomous variable (Yes/No) "Used marijuana 1 or more times in the past month" was used. The original item presented to the respondents was worded as "How many times have you smoked marijuana in the previous month," and was coded as 0 times, 1–2, 3–9, 10–19, 20–39, or 40 or more times.

Within the content area of substance use, selected items reflected both substance use across one's lifetime as well use within the past few weeks. Based on previous literature regarding behaviors that might reflect adolescents at-risk, items related to cigaret, alcohol, marijuana, cocaine, heroin, ecstasy, methamphetamines, and prescription drug use and the relationship to health behaviors were included (Wickrama et al., 1999; Safron et al., 2001; Hamilton, 2009;

Table 1 | Items from the YRBS used in the current study.

Item*
Smoked 1+ cigaret in the past week
Used snuff or dip 1+ days in the past 30 days
Had 1 or more drinks in the past 30 days
Used marijuana one or more times in the past month
Used cocaine one or more times in lifetime
Sniffed glue one or more times in lifetime
Used heroin one or more times in lifetime
Used meth one or more times in lifetime
Used ecstasy one or more time in lifetime
Offered or sold drugs at school
Taken prescription drugs w/o prescription
Had sex before age 13
Had sex with more than one person in the past 3 months
Had more than four sexual partners

All items were responded to as Yes or No.

Landis et al., 2009; Grana et al., 2010), whereas drugs related to possible performance enhancement such as steroids were excluded from the analysis. With respect to risky sexual behaviors, items related to age of onset of sexual behavior, number of partners in one's lifetime, and number of partners within the last 3 months were selected. Previous findings indicate that adolescents who are at greater risk for a variety of psycho-social problems are more likely to have had sexual experiences and experiences with more partners than those without (Brown et al., 2010). This is an important question because of the tendency for many individuals to be discrete in reporting sexual histories. If different rates are found for different classes then it may be possible to more efficiently ask about sexual behavior in future research efforts.

PARTICIPANTS

Sampling for the YRBS was done using a three-stage cluster sample stratified by ethnic concentration and Metropolitan statistical area (MSA). Thus, each individual was assigned a sampling weight, which was used in all analyses for this study. In 2009 the CDC sampled students from 196 schools that included public, private, and religious schools to participate in the YRBS effort. Of this sample 158 schools participated, resulting in a response rate of 81%, with 88% (16,410 out of 18,573) of individual students responding. Within the school, data was collected through the use of a random sampling approach to identify classes and then students for participation (Centers for Disease Control, 2009). The size of the dataset ($N = 16,410$) and the strategy for data collection allows for a meaningful analysis of minority groups such as Native Americans, and Asian American students that often are not included in research studies. The dataset also includes information for rural, suburban, and urban communities and students. As a result the findings of the analysis are likely to have a high degree of external validity and a good fit to the population of students across the United States. The sample used in this analysis consisted of the 16,410 individuals who responded to the YRBS. Descriptive statistics for sample demographics appear in **Table 2**. A total of 52.2% of the sample

was male, while 58.7% was Caucasian. The largest percentage of students was in the 9th grade, and the smallest was in the 12th. On average the members of this sample were 16.03 years old with a SD of 1.23.

DATA ANALYSIS

In order to identify subgroups in the population based on responses to the risky behavior items described above, a multilevel mix2PL model was used (Vermunt, 2003). Data were clustered by school which the students attended. It should be noted that this approach is only one of several that could have been applied in this situation. For example, it would have been possible to use a LCA approach in order to identify population subgroups based on their risky behaviors. Likewise, an IRT model could also have been employed in order to obtain parameter estimates for those items of particular interest. We elected to use the 2PL mixture model in order to satisfy the two primary goals of this study: (1) Identify subgroups at-risk for potentially self destructive behavior and (2) Determine how these specific items on the YRBS might perform differently for these subgroups. This latter goal would be of particular interest to psychologists and others who make use of the instrument to help identify at-risk youth. Initially, a multidimensional mix2PL model was also used with the data, but the fit indices indicated that it was not as appropriate as the unidimensional model that we eventually settled on.

Parameters were estimated using MLE with robust SE with MPlus 5.2 software (Muthén and Muthén, 2008), with model identification coming from setting class latent trait means to 0. As mentioned previously, in order to avoid the local maxima problem, each model was fitted with 10 random starting values (Rost, 1991). Models were estimated for 2, 3, 4, 5, and 6 latent classes. In order to determine the optimal number of latent classes, the sample size adjusted BIC (SBIC) was used, where smaller values indicate better model fit to the data (Yang, 2006; Tofighi and Enders, 2007). In addition, per recommendations in the literature (e.g., Bauer and Curran, 2004) substantive considerations regarding the coherence of the latent classes were also taken into consideration when deciding on the optimal solution. The latent classes were characterized using both the item parameter estimates that are a part of the mixIRT framework, as well as by descriptive statistics of demographic and behavior variables, and a multivariate analysis of variance (MANOVA) and discriminant analysis (DA) comparing the latent classes on the sums of risky sexual and substance use behaviors engaged in.

RESULTS

MIXTURE IRT LATENT CLASS MEMBERSHIP

The SBIC values for each number of latent classes appear in **Table 3**. Based on these indices, it appears that the 4 class solution provided the best relative fit to the YRBS data. In addition, an examination of the 3, 4, and 5 group solutions based on the substantive meanings of the groups also supported the 4 class solution. When 3 classes were retained, groups 1 and 2 of the 4 class solution were joined together, although as will be discussed below they demonstrated some fundamentally different item responses and demographic patterns. For the 5 class solution, group 4 in the 4 class result was divided into two groups, though substantively

Table 2 | Descriptive statistics for total YRBS sample.

Variable	Class 1	Class 2	Class 3	Class 4	Total
GENDER					
Female	981 (46.0%)	1773 (47.3%)	255 (48.7%)	4822 (48.2%)	7831 (47.7%)
Male	1152 (54.0%)	1976 (52.7%)	268 (51.3%)	5183 (51.8%)	8579 (52.3%)
ETHNICITY					
Caucasian	1058 (49.6%)	2287 (61.0%)	388 (74.2%)	5863 (58.6%)	9596 (58.5%)
African-American	512 (24.0%)	487 (13.0%)	54 (10.3%)	1341 (13.4%)	2394 (14.6%)
Other	563 (26.4%)	975 (26.0%)	81 (15.5%)	2801 (28.0%)	4420 (26.9%)
GRADE					
9	491 (23.0%)	859 (23.1%)	130 (24.3%)	2843 (28.0%)	4323 (26.3%)
10	467 (21.9%)	961 (25.5%)	144 (28.1%)	2665 (26.2%)	4237 (25.8%)
11	570 (26.7%)	990 (26.5%)	113 (21.1%)	2234 (21.9%)	3907 (23.8%)
12	605 (28.4%)	939 (24.9%)	136 (25.2%)	2263 (22.2%)	3943 (24.1%)
Age mean (SD)	16.70 (1.19)	16.21 (1.13)	16.49 (1.19)	15.93 (1.24)	16.03 (1.23)
Sexual behaviors mean (SD)	1.88 (0.93)	0.72 (0.65)	1.40 (1.21)	0.36 (0.69)	0.78 (0.97)
Substance use mean (SD)	2.29 (2.37)	2.74 (2.17)	5.56 (2.29)	0.52 (2.02)	1.48 (2.23)
POSTERIOR PROBABILITY					
1	0.884	0.057	0.179	0.012	
2	0.057	0.756	0.066	0.120	
3	0.179	0.123	0.669	0.030	
4	0.012	0.140	0.007	0.840	
Total	2133 (13.0%)	3749 (22.9%)	523 (3.2%)	10,005 (61.0%)	

Table 3 | Size adjusted BIC values for 2, 3, 4, 5, and 6 latent classes.

Latent classes	SBIC
2	158438.172
3	142049.551
4	141702.480
5	141777.782
6	141987.966

their item parameter values were very similar. Thus, given that the 4 class solution yielded the lowest SBIC value and substantively was the most coherent, it was selected for further examination in the current study.

The number and proportion of each of the four classes appear in **Table 2**, along with the posterior probabilities of group membership. These latter values suggest that there is good group separation, with relatively little overlap across the four classes. Latent class 4 was by far the largest, accounting for more than 60% of the total sample, while class 3 was the smallest, with less than 5% of the sample. Demographic information by latent class also appears in **Table 2**. These results indicate that Class 1 had a somewhat greater representation of males than was present in the overall sample, while in the other classes the genders appeared in rates comparable to that of the sample as a whole. Classes 3 and 4 consisted of more students in grades 9 and 10, while class 1 was made up of a majority of 11 and 12th graders, and class 2 displayed a somewhat more equal distribution of the four grade levels. In terms of ethnicity, classes 2 and 3 had higher percentages of Caucasian subjects than was typical for the entire sample, while class 1 was disproportionately

more African-American. Latent classes 1, 2, and 3 were similar with respect to mean age, with class 4 being between 0.3 and 0.7 years younger than the others, on average.

In addition to characterizing the latent classes based on their demographic characteristics, it is also useful to compare them in terms of their behaviors, as reflected in responses to items on the YRBS. For each subject in the sample the number of items endorsed for each of the sexually risky (maximum of 3) and substance abuse (maximum of 12) behaviors that were used in the mixture IRT analysis were summed to create risky behavior scores. A higher number of such endorsements indicates that the adolescent engaged in more of the risky behaviors. For each latent class, the mean number of risky sexual and substance use behaviors was then calculated, and MANOVA and DA were used to explore whether and how the latent classes differed with respect to these two behavior variables. The assumptions of normality and homogeneous covariance matrices were assessed and found to be satisfied. Two significant discriminant functions were identified ($p < 0.001$ for both). Absolute values of the structure coefficients (**Table 4**) greater than 0.32 were taken to mean that the behavioral variable was important in differentiating the latent classes for the discriminant function (Tabachnick and Fidell, 2007). Thus, it appears that function 1 was primarily associated with sexually risky behaviors, but also was influenced by substance use, while function 2 was only associated with substance use. An examination of the mean number of risky behaviors that were endorsed (**Table 2**) shows that latent class 3 had the highest mean for substance use, while class 1 had the highest mean for risky sexual behaviors. On the other hand, class 4 had the lowest means for both sets of items, indicating that they endorsed these behaviors

the least often. Finally, class 2 endorsed the second largest number of substance use behaviors, while class 3 endorsed the second most sexual behaviors.

MIXTURE IRT ITEM PARAMETER ESTIMATES

As discussed previously in the literature (Lazarsfeld and Henry, 1968; Rost, 1990; Cohen and Bolt, 2005) an examination of item parameter values can further aid in understanding the nature of the latent classes identified in the mixture IRT analysis by identifying items that each found disproportionately easy (or difficult) to endorse, and which were disproportionately better (or worse) at differentiating individuals based on the latent trait being measured, in this case the propensity to engage in risky behavior. In this study, item "difficulty" reflects the propensity of an individual to endorse having engaged in a specific behavior (e.g., smoking within the last 30 days), while discrimination provides information regarding the ability of a specific item to differentiate individuals on the latent trait being measured, which in this case is propensity for engaging in risky behaviors. Therefore, items with relatively high discrimination values will be more effective at distinguishing among individuals with different propensities to engage in risky behavior. Latent class differences on the item difficulty and discrimination parameter values can thus provide insights into how the groups differ from one another, and also give specific information to mental health professionals regarding which behaviors might be most effective for identifying at-risk adolescents who fit a specific typology (i.e., latent class) seen in the population. The item difficulty and discrimination parameter estimates and SE for the mix2PL model appear in **Tables 5 and 6**, respectively.

Item difficulty

Higher values for item difficulty parameter estimates indicate that required a greater level of the latent trait being measured (i.e., must be more at-risk) in order to endorse the behavior. For example, members of class 1 found the items "Had sex with four or more people in lifetime" and "Had sex with one or more people in the last 3 months" easier to endorse than did members of the other latent classes. This means that they did not have to be particularly high on the at-risk latent trait in order to respond that they have had sex with four or more people in their lifetime or that they have had sex with one or more people in the last 30 days. In addition, group 1 also found the item "Had one or more drinks of alcohol in the last 30 days" easier to endorse than did members of groups 3 or 4, but found the item "Had sex before the age of 13" more difficult to endorse than did members of any of the other latent classes. In contrast, class 3 found the item "Had sex prior to the age of 13" much easier to endorse than did the other groups. In addition, these adolescents also had lower difficulty parameter estimates than the other classes for most of the substance use items with the exception of "Have you had one or more drinks of alcohol in the last 30 days." As noted above, this latent class also had the highest mean value for substance use item endorsements. Finally, members of latent class 3 had higher difficulty estimates than the other groups, except four, on the item regarding alcohol consumption in the last 30 days, suggesting that they found this particular substance use behavior less easy to endorse. In contrast, latent class 2 found the items regarding smoking and drinking alcohol in the last 30 days easier to endorse than did any of the other groups. However, with respect to other substances, their item difficulty values were only somewhat smaller than those of class 1, and larger than those of class 3. Thus, latent class 2 can be characterized as finding it relatively easy to endorse smoking tobacco and drinking alcohol, but not using other substances. In addition, latent class 2 found the items regarding having sex prior to age 13 or with four or more partners in their lifetime particularly easy to endorse. Finally, adolescents in latent class 4 had the highest difficulty parameter estimates for all of the risk behavior items included in this analysis. This result

Table 4 | Structure coefficients for discriminant analysis comparing cluster means on number of risky sexual and substance use behaviors.

Variable	Function 1	Function 2
Sexual behaviors	0.997	−0.082
Substance use behaviors	0.536	0.844

Table 5 | Item difficulty parameter estimates (SE) by latent class.

Item	Class 1	Class 2	Class 3	Class 4
Had sex before 13	4.817 (2.492)	1.188 (0.143)	−2.522 (1.003)	3.824 (0.405)
Had sex with 4+	−7.447 (2.688)	0.099 (0.267)	2.306 (0.017)	3.054 (0.282)
Had sex with 1+ month	−5.507 (2.281)	−0.835 (0.208)	−2.002 (0.459)	2.819 (0.429)
Smoked 1+ last 30 days	0.268 (0.180)	−1.022 (0.447)	−0.547 (0.139)	3.177 (0.345)
snuff 1+ last 30 days	1.743 (0.392)	−0.001 (0.459)	−2.569 (0.298)	5.535 (1.244)
Drink 1+ last 30 days	−0.699 (0.103)	−1.525 (0.391)	1.248 (0.516)	1.847 (0.185)
Marijuana 1+ last 30 days	0.022 (0.120)	0.206 (0.140)	−0.262 (0.173)	3.025 (0.281)
cocaine 1+ in life	1.025 (0.067)	0.729 (0.097)	−1.627 (0.121)	2.059 (0.080)
Sniff glue 1+ in life	1.201 (0.153)	0.747 (0.023)	−1.766 (0.119)	2.037 (0.086)
Heroin 1+ in life	1.444 (0.103)	1.348 (0.023)	−2.398 (0.211)	2.422 (0.076)
Meth 1+ in life	1.299 (0.075)	0.867 (0.270)	−2.085 (0.155)	2.207 (0.066)
Ecstasy 1+ in life	1.056 (0.137)	0.681 (0.019)	−1.725 (0.123)	2.233 (0.077)
Offered/sold drugs school	0.713 (0.069)	0.626 (0.105)	−1.048 (0.104)	1.948 (0.132)
LSD 1+ times	1.154 (0.084)	0.352 (0.682)	−1.536 (0.114)	2.212 (0.079)
Prescription drugs	0.546 (0.132)	0.001 (0.001)	−0.851 (0.102)	1.868 (0.102)

means that members of this group had to have higher values on the at-risk latent trait than those in other latent classes in order to report having engaged in these risky behaviors. Note that this latent class also had the lowest mean number of such behaviors endorsed.

Using the different patterns in the item difficulty parameter estimates across the latent classes, we can characterize the four groups in terms of their risky behaviors. Members of group 1, for example, found it relatively easier than members of the other classes to report engaging in risky sexual behaviors, and alcohol consumption, but more difficult than some of the other groups to endorse items regarding the use of other drugs. Thus, this group might be thought of as at heightened risk for sexual activity. In contrast, adolescents in class 3 found items dealing with drugs other than tobacco and alcohol easier to endorse than did members of any of the other groups, but they found the risky sexual behaviors somewhat more difficult than several other classes. This group, then, would be seen as more at-risk for using drugs than those in other classes. Members of latent class 2 were characterized as finding tobacco and alcohol use disproportionately easy as compared to the other groups, though they found items pertaining to the use of other drugs relatively more difficult. In addition, this latent class found risky sexual behavior somewhat more difficult to endorse as well. Finally, latent class 4 found all of the risk behavior items more difficult to endorse than members of the other three groups, suggesting that they are the least likely to engage in any of these risky behaviors at a given level of the at-risk latent trait. In short, group 1 can perhaps be characterized as more willing to engage in risky sexual behavior and alcohol consumption, group 2 as more willing to smoke tobacco and drink alcohol, but avoid other drugs, group 3 as more willing to have sex at a very young age, and use a variety of drugs other than alcohol and tobacco, and group 4 as finding it the most difficult to endorse participation in any of the risky behaviors.

Item discrimination

As described briefly above, discrimination refers to ability of an item to differentiate between individuals with different values of the latent trait being measured (De Ayala, 2009). Therefore, with

the mix2PL model used in this study it is possible to ascertain for which groups a given item is particularly effective at differentiating adolescents who have a greater propensity to engage in risky behaviors from those who are at lower such risk. For latent class 1, results in **Table 6** show that the item discrimination values were reasonably large, with the exception of items asking about risky sexual behavior. Recall that based on the item difficulty estimates, class 1 found risky sexual behavior items to be relatively easy to endorse. These results suggest that items regarding substance use would be potentially effective for identifying at-risk adolescents in class 1, but that items pertaining to sexual behavior would not. A similar pattern was in evidence for class 3, for whom the items regarding sex before the age of 13 and sex with one or more people in the last month were poor discriminators, though having sex with four or more people in their lifetime was a very good discriminator for this class. In contrast, only the item pertaining to snuff use one or more times in the last 30 days had a low discrimination value for group 2. Finally, for group 4, none of the items had extremely low discrimination values, although the items regarding having had sex with one or more people in the last 30 days, and using snuff had somewhat lower values than did the others.

The ability to obtain class specific item discrimination parameter estimates is potentially very useful in this context, because it allows mental health professionals to focus on behaviors that might be most effective for identifying those particularly at-risk within specific typologies of risk. As an example of using discrimination values in this way, consider that among those in latent class 1 the items "Used heroin one or more times in your life," "Used ecstasy one or more times in your life," and "Used LSD one or more times in your life" displayed the highest discrimination estimates, and would thus be most effective at differentiating adolescents based on their overall propensity for engaging in risky behaviors. On the other hand, none of the sexual behavior items would be useful in this regard. For class 3, which found endorsing most of the substance use behaviors relatively easy, the items asking about cocaine, heroin, methamphetamine, and LSD

Table 6 | Item discrimination parameter estimates (SE) by latent class.

Item	Class 1	Class 2	Class 3	Class 4
Had sex before 13	0.096 (0.052)	1.735 (0.671)	0.083 (0.198)	0.689 (0.101)
Had sex with 4+	0.025 (0.384)	1.549 (0.617)	2.812 (0.189)	1.203 (0.182)
Had sex with 1+ month	0.236 (0.117)	1.146 (0.526)	0.057 (0.063)	0.351 (0.067)
Smoked 1+ last 30 days	1.207 (0.271)	0.568 (0.407)	1.009 (0.119)	0.928 (0.130)
Snuff 1+ last 30 days	0.845 (0.140)	0.279 (0.205)	0.547 (0.083)	0.483 (0.126)
Drink 1+ last 30 days	0.923 (0.103)	1.339 (0.490)	0.629 (0.150)	0.632 (0.070)
Marijuana 1+ last 30 days	0.918 (0.114)	2.105 (0.670)	0.836 (0.118)	1.062 (0.187)
Cocaine 1+ in life	1.857 (0.240)	3.198 (1.229)	1.654 (0.201)	2.459 (0.204)
Sniff glue 1+ in life	1.075 (0.145)	2.106 (0.770)	0.722 (0.067)	0.978 (0.060)
Heroin 1+ in life	3.415 (0.642)	4.512 (0.407)	2.133 (0.470)	3.304 (0.684)
Meth 1+ in life	2.477 (0.302)	6.163 (0.835)	1.765 (0.210)	2.853 (0.271)
Ecstasy 1+ in life	1.606 (0.163)	2.189 (0.668)	1.359 (0.151)	2.752 (0.352)
Offered/sold drugs school	0.432 (0.062)	1.918 (0.859)	0.529 (0.058)	0.547 (0.039)
LSD 1+ times	2.151 (0.246)	6.212 (0.975)	1.643 (0.183)	2.359 (0.237)
Prescription drugs	1.029 (0.110)	3.130 (0.637)	1.150 (0.118)	1.009 (0.067)

use proved to be the most effective at differentiating individuals based on their at-risk status, as was the item "Had sex with four or more people in your lifetime." For class 2, who found it easier to endorse items about having sex with one or more people, smoking tobacco, and drinking alcohol in the last 30 days, several of the substance use items proved to be strong discriminators of at-risk status, including the use of marijuana in the last 30 days, the use of cocaine, glue sniffing, heroin, methamphetamine, ecstasy, LSD, and prescription drugs without a prescription. Finally, for group 4, which had high difficulty parameter estimates for all of the risky behavior items, the use of cocaine, heroin, methamphetamine, ecstasy, and LSD were the best discriminators of individuals based on their at-risk latent trait, though in general all of the items appear to have been reasonably useful in this regard.

DISCUSSION

The goals of this study were twofold. First, we wanted to demonstrate the utility and power of the mixIRT modeling framework beyond its uses in achievement testing and personality assessment, to the realm of behavioral measures such as the YRBS. The ability to simultaneously identify latent classes and to estimate separate IRT models for these classes can provide researchers with greater insights into differential behavioral patterns in the population and which behaviors might be best at differentiating people based on the latent trait being measured. Secondly, the results of the study provide information specific to the YRBS on the identification of individuals at-risk for engaging in risky sexual and substance use behaviors.

The mixIRT analysis of the risky sexual and substance use behavior items on the YRBS revealed the presence of 4 latent classes among adolescents. Class 1 was somewhat older and more male than the other classes, and was also more likely to have engaged in risky sexual behaviors. Members of classes 2 and 3 were more likely to be Caucasian, and class 2 found it easier to endorse items indicating use of alcohol and cigarets, while class 3 found it easier to endorse items reflective of a wide range of substance use. Finally, members of latent class 4 tended to be somewhat younger than those in other classes, and displayed the least inclination to endorse any of the risky behavior items included in this study. In addition, the results of the mix2PL analysis reveal which items are most useful for differentiating high and low risk individuals within each latent class. Such results can prove particularly useful to clinical practitioners as well as instrument developers, allowing for the use of class specific behaviors in identifying potentially at-risk individuals. For example, among those in latent class 1, items asking about the use of drugs such as cocaine, glue sniffing, heroin, Methamphetamine, and LSD had higher discrimination values and thus would be quite useful in differentiating individuals at greater risk. Conversely, the items regarding risky sexual behavior would serve as good discriminators for at-risk youth in latent class 2. More generally, we can conclude from these results that different items will be more useful for differentiating at-risk adolescents with different behavior profiles (i.e., different latent classes). In addition, it is possible to discern which typology an individual is likely to fall into, based on their pattern of item endorsements. Those who have engaged in more sexual behavior but not as much substance use, for example, are more likely to

be in latent class 1, while those who report using a wider variety of substances are more likely to be in latent class 3, and so on. Knowing this information would allow clinicians to focus on specific types of behaviors in order to identify those at greatest risk within the specific typologies.

In addition to the content specific goal of identifying typologies of adolescents at-risk for engaging in risky sexual and substance use behaviors, the other major focus of this manuscript was to demonstrate the power and utility of the mixIRT modeling framework. This methodology is still fairly new, and therefore its potential for application in a wide variety of contexts has not yet been fully realized. It is hoped that this study makes manifest the great potential mixIRT has in the broader psychology literature as a tool for identifying psychological typologies within the population, as well as for finding measurement scale items that can be used to differentiate individuals from these various typologies. Other commonly used methods for addressing such research scenarios are limited to either identifying subgroups (e.g., LCA, cluster analysis) or modeling item response profiles for a single, or several known groups (e.g., IRT, factor analysis). Each of these approaches has limitations that mixIRT is able to overcome and in so doing provide researchers with a richer tapestry of research results in the form of latent class identification and item parameter estimation. It is hoped that this study has demonstrated, using one example from health psychology, how these models can be used.

FUTURE DIRECTIONS FOR RESEARCH

There are a number of directions for future research in understanding the performance of, and applying the mixIRT model. The current study made use only of dichotomously scored items (behavior present or behavior absent). However, mixIRT models are able to accommodate ordinal data as well, including responses to Likert items. Future research in psychology should more widely employ models for ordinal data, including the mixture versions of the Graded Response Model and the Partial Credit Model, to examine severity and frequency of feelings or behaviors (von Davier and Rost, 1995; von Davier and Yamamoto, 2004). In addition, using mixIRT models it is possible to obtain estimates of the latent trait for each member of the sample. Thus, in a study such as this one, it would be possible to obtain estimates for this propensity, and to use them for identifying specific individuals with very high latent trait estimates who might be at the greatest risk. There remain questions regarding the optimal methods for parameter estimation (MLE versus Bayesian), as well as the accuracy of these estimates under a variety of real world data conditions. In the current study, only the 2PL mixture model was selected for use, based on our beliefs about the nature of the items. However, in some instances of practice model selection can be made based on ascertaining and comparing the goodness of fit of the models using tools such as information criteria such as the Bayesian information criterion (BIC) and the Akaike information criterion (AIC). In the context of mixture IRT models, it is not clear which of these approaches to model selection might be optimal. A number of these methodological issues should be investigated using Monte Carlo simulations.

The results of this paper and the methodology that it demonstrates may be helpful in contributing to our understanding of the development of at-risk behaviors in additional ways. While

in the current study data from individuals across the entire high school period was analyzed, it is also possible to, in a similar fashion to the work done with cognitive and achievement measures, evaluate the scales for specific age cohorts. In this way, age specific developmental differences can be removed from the analysis, thus potentially refining the measurement of the latent trait of interest. As a result, questionnaires of psychopathology and at-risk behavior may be developed that suggest whether an individual is engaging in age appropriate types of risky or psychopathological behaviors.

REFERENCES

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

Bekman, N. M., Cummins, K., and Brown, S. A. (2010). Affective and personality risk and cognitive mediators of initial adolescent alcohol use. *J. Stud. Alcohol Drugs* 71, 570–580.

Berk, L. E. (2008). *Infants, Children, and Adolescence*, 6th Edn. Boston, MA: Pearson Education, Inc.

Bolt, D. M., Cohen, A. S., and Wollack, J. A. (2001). A mixture item response model for multiple-choice data. *J. Educ. Behav. Stat.* 26, 381–409.

Bolt, D. M., Cohen, A. S., and Wollack, J. A. (2002). Item parameter estimation under conditions of test speededness: application of a mixture Rasch model with ordinal constraints. *J. Educ. Meas.* 39, 331–348.

Brown, L. K., Hadley, W., Stewart, A., Lescano, C., Whiteley, L., Donenberg, G., and DiClemente, R. (2010). Psychiatric disorders and sexual risk among adolescents in mental health treatment. *J. Consult. Clin. Psychol.* 78, 590–597.

Centers for Disease Control. (2009). *Youth Risk Behavior Survey: 2009 National YRBS Data Users Manual.*

Cohen, A. S., and Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *J. Educ. Meas.* 42, 133–148.

Coley, R. L., Votruba-Drazl, E., and Schindler, H. S. (2009). Fathers' and mothers' parenting predicting and responding to adolescent sexual risk behaviors. *Child Dev.* 80, 808–827.

De Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* New York: The Guilford Press.

De Ayala, R. J., Kim, S.-H., Stapleton, L. M., and Dayton, C. M. (2002). Differential item functioning: a mixture distribution conceptualization. *Int. J. Test.* 2, 243–276.

Eid, M. (1997). "Happiness and satisfaction: an application of a latent state-trait model for ordinal variables," in *Applications of Latent Trait and Latent Class Models in the Social Sciences*, eds J. Rost and R. Langeheine. Available at: http://www.ipn.uni-kiel.de/aktuell/buecher/rostbuch/ltlc.htm [Retrieved February 4, 2011].

Eid, M., and Zickar, M. (2007). "Detecting response styles and faking in personality and organizational assessment by mixed Rasch models," in *Multivariate and Mixture Distribution Rasch Models*, eds M. von Davier and C. Cartensen (New York: Springer), 255–270.

Fossati, A., Maffei, C., Bagnato, M., Battaglia, M., Donati, D., Donini, M., Fiorilli, M., and Novella, L. (2001). Latent class analysis of DSM-IV schizotypal personality disorder criteria in psychiatric patients. *Schizophr. Bull.* 27, 59–71.

Frick, U., Rehm, J., and Thien, U. (1997). "On the latent structure of the Beck depression inventory (BDI): using the "Somatic" subscale to evaluate a clinical trial," in *Applications of Latent Trait and Latent Class Models in the Social Sciences*, eds J. Rost and R. Langeheine. Available at: http://www.ipn.uni-kiel.de/aktuell/buecher/rostbuch/ltlc.htm [Retrieved February 4, 2011].

Grana, R. A., Black, D., Sun, P., Rohrbach, L. A., Gunning, M., and Sussman, S. (2010). School disrepair and substance use among regular and alternative high school students. *J. Sch. Health* 80, 387–393.

Hamilton, G. J. (2009). Prescription drug abuse. *Psychol. Sch.* 46, 892–898.

Hamrin, V., McCarthy, E. M., and Tyson, V. (2010). Pediatric psychotropic medication initiation and adherence: a literature review based on social exchange theory. *J. Child Adolesc. Psychiatr. Nurs.* 23, 151–172.

Harley, M., Kelleher, I., Clarke, M., Lynch, F., Arsenault, L., Connor, D., Fitzpatrick, C., and Cannon, M. (2010). Cannabis use and childhood trauma interact additively to increase the risk of psychotic symptoms in adolescence. *Psychol. Med.* 40, 1627–1634.

Hirschi, T. (1969). *Causes of Delinquency.* Berkeley: University of California Press.

Junker, B. W. (1999). Some statistical models and computational methods that may be useful for cognitively-relevant assessment. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.7053&rep=rep1&type=pd [Retrieved November 8, 2010].

Landis, A. M., Parker, K. P., and Dunbar, S. B. (2009). Sleep, hunger, satiety, food cravings, and caloric intake in adolescence. *J. Nurs. Scholarsh.* 41, 115–123.

Lazarsfeld, P. F., and Henry, N. W. (1968). *Latent Structure Analysis.* Boston: Houghton Mifflin.

Lee, C.-Y. S., Winters, K. C., and Wall, M. M. (2010). Trajectories of substance use disorders in youth: identifying and predicting group memberships. *J. Child Adolesc. Subst. Abuse* 19, 135–157.

Li, F., Cohen, A. S., Kim, S.-H., and Cho, S.-J. (2009). Model selection methods for mixture dichotomous IRT models. *Appl. Psychol. Meas.* 33, 353–373.

Maij-de Meij, A. M., Kelderman, H., and van der Flier, H. (2008). Fitting a mixture item response model to personality questionnaire data: characterizing latent classes and investigating possibilities for improving prediction. *Appl. Psychol. Meas.* 32, 611–631.

McConaughy, S. H., and Ritter, D. R. (2008). "Best practices in multi-method assessment of emotional and behavioral disorders," in *Best Practices in School Psychology V*, eds A. Thomas and J. Grimes (Bethesda, MD: National Association of School Psychologists), 697–715.

Mislevy, R. J., and Verhelst, N. D. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika* 55, 195–215.

Muthén, L. K., and Muthén, B. O. (2008). *MPlus Software, Version 5.2.* Los Angeles, CA: MPlus.

Prado, G., Huang, S., Maldonado-Molina, M., Bandiera, F., Schwartz, S. J., de la Vega, P., Brown, S. H., and Pantin, H. (2010). An empricial test of ecodevelopmental theory in predicting HIV risk behaviors among Hispanic youth. *Health Educ. Behav.* 37, 97–114.

Reininger, B. M., Evans, A. E., Griffin, S. F., Sanderson, M., Vincent, M. L., Valois, R., and Parra-Medina, D. (2005). Predicting adolescent risk behaviors based on an ecological framework and assets. *Am. J. Health Behav.* 29, 150–161.

Rost, J. (1990). Rasch models in latent classes: an integration of two approaches to item analysis. *Appl. Psychol. Meas.* 14, 271–282.

Rost, J. (1991). A logistic mixture distribution model for polychotomous item responses. *Br. J. Math. Stat. Psychol.* 44, 75–92.

Safron, D. J., Schulenberg, J. E., and Bachman, J. G. (2001). Part-time work and hurried Adolescence: the links among work intensity, social activities, health behaviors, and substance use. *J. Health Soc. Behav.* 42, 425–449.

Sattler, J. D., and Hoge, R. D. (2008). *Assessment of Children: Behavioral, Social, and Clinical Foundations*, 5th Edn. San Diego, CA: Sattler Publishing, Inc.

Shih, R. A., Miles, J. N. V., Tucker, J. S., Zhou, A. J., and D'Amico, E. J. (2010). Racial/ethnic differences in adolescent substance use: mediation by individual, family, and school factors. *J. Stud. Alcohol Drugs* 71, 640–651.

Smit, A., Kelderman, H., and van der Flier, H. (2003). Latent trait latent class analysis of an Eysenck Personality Questionnaire. *Methods Psychol. Res. Online* 8, 23–50.

Sorenson, A. M., Brownfield, D., and Jensen, G. F. (1997). "Constructing comparable measures of delinquency by gender: a theoretical and empirical inquiry," in *Applications of Latent Trait and Latent Class Models in the Social Sciences*, eds J. Rost and R. Langeheine. Available at: http://www.ipn.uni-kiel.de/aktuell/buecher/rostbuch/ltlc.htm [Retrieved February 4, 2011].

Tabachnick, B. G., and Fidell, L. S. (2007). *Using Multivariate Statistics.* Boston, MA: Pearson Education, Inc.

Tofighi, D., and Enders, C. K. (2007). "Identifying the correct number of classes in a growth mixture models," in *Advances in Latent Variable Mixture Models*, eds G. R. Hancock and K. M. Samuelsen (Greenwich, CT: Information Age), 317–341.

Uebersax, J. S. (1997). "Analysis of student problem behaviors with latent trait, latent class, and related probit mixture models," in *Applications of Latent Trait and Latent Class Models in the Social Sciences*, eds J. Rost and R. Langeheine. Available at: http://www.ipn.uni-kiel.de/aktuell/buecher/rostbuch/ltlc.htm [Retrieved February 4, 2011].

van der Heijden, P., 't Hart, H., and Dessens, J. (1997). "A parametric bootstrap procedure to perform statistical tests in a LCA of anti-social behavior," in *Applications of Latent Trait and Latent Class Models in the Social Sciences*, eds J. Rost and R. Langeheine. Available at: http://www.ipn.uni-kiel.de/aktuell/buecher/rostbuch/ltlc.htm [Retrieved February 4, 2011].

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

Vierhaus, M., and Lohaus, A. (2008). Children and parents as informants of emotional and behavioral problems

predicting female and male adolescent risk behavior: a longitudinal cross-informant study. *J. Youth Adolesc.* 37, 211–224.

von Davier, M., and Rost, J. (1995). "Polytomous mixed Rasch models," in *Rasch Models: Foundations, Recent Developments and Applications*, eds G. H. Fischer and I. W. Molenaar (New York: Springer), 371–379.

von Davier, M., and Rost, J. (2007). "Mixture distribution item response models," in *Handbook of statistics, 26: Psychometrics*, eds C. R. Rao and S. Sinharay (Amsterdam: Elsevier), 643–661.

von Davier, M., and Yamamoto, K. (2004). Partially observed mixture models of IRT models: an extension of the generalized partial credit model. *Appl. Psychol. Meas.* 28, 389–406.

Wickrama, K. A. S., Conger, R. D., Wallace, L. E., and Elder, G. H. (1999). The intergenerational transmission of health-risk behaviors: adolescent lifestyles and gender moderating effects. *J. Health Soc. Behav.* 40, 258–272.

Yamamoto, K. (1987). *A Model That Combines IRT and Latent Class Models.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Champaign, IL.

Yang, C. (2006). Evaluating latent class analyses in qualitative phenotype identification. *Comput. Stat. Data Anal.* 50, 1090–1104.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.