An Empirical Comparison of Three Psychometric Measures of Adolescent Substance Use

Miriam Brinberg

The University of North Carolina at Chapel Hill

Spring 2015

A thesis presented to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the Bachelor of Science degree with Honors in Psychology.

Advisor___________________________
Dr. Patrick Curran

Reader___________________________
Dr. Mitchell Prinstein

Reader___________________________
Dr. Kathleen Gates

Acknowledgements

I would first like to thank my advisor, Dr. Patrick Curran. He successfully guided me through completing my honors thesis, which has been a wonderful challenge and incredible learning experience, and he also provided tremendous amounts of time and advice about the field of quantitative psychology and graduate school applications. During both my junior year as a student within his class and during my senior year as his honors student, I appreciate the time and effort he took to meet with me and help me understand the material. I certainly would not have such a great project (and gotten into graduate school) without his patience and guidance!

Secondly, I would like to thank the members of my committee, Dr. Mitchell Prinstein and Dr. Katie Gates. I began my research experience in Dr. Prinstein's "Peer Relations Lab" where I had the opportunity to learn through experience in the lab working with the data and through our weekly lab discussions; I also appreciate his professional development advice throughout my college career. Dr. Gates provided much encouragement in her "General Linear Models" class. While the material could be difficult, her caring attitude and patience truly made the class enjoyable and reinforced my interest in the field of quantitative psychology.

Finally, I would like to thank my family. My parents and brother were always available to bounce ideas off of, to listen to my concerns throughout the process, to read through drafts, and for being the loving and supportive people that they are.

Abstract

Adolescent substance use is common and often accompanied by many negative consequences. The psychometric assessment of substance use, however, is highly varied and may misrepresent actual substance use. Given the importance of understanding adolescent substance use, this study examined three competing measurement models of substance use and their differential predictability of academic achievement and internalizing and externalizing symptomatology. Participants were drawn from the public-use Add Health data set ($N = 5{,}857$). The three measurement models were determined from seven items that assessed the lifetime use of cigarettes, chewing tobacco, alcohol, marijuana, cocaine, inhalants, and other illicit drugs. The first measurement model (lifetime use) grouped participants as abstainers (never used a drug) and users (used at least one drug). The second measurement model (proportion) refined substance users by accounting for the proportion of the seven drugs they endorsed. The final measurement model further refined substance users by creating factor scores through confirmatory factor analysis that allow each item to be differentially weighted as a function of severity. Nine separate regression analyses were estimated in which age, race, gender, and substance use measurement model predicted either academic achievement or externalizing or internalizing symptomatology. These analyses indicated that all three substance use measures were significant predictors of each outcome measure, but the proportion and factor score models accounted for substantially more of the variance within each model. These results suggest the proportion or factor score models would be better predictors of substance use, although the choice of model should be based on relevant theory.

An Empirical Comparison of Three Psychometric Measures of Adolescent Substance Use

Adolescent substance use has many adverse consequences, including future addiction, unintended pregnancies, mental health issues, injuries, and even death, with 68% of adolescents reporting trying alcohol by the end of high school (The National Center on Addiction Substance Abuse at Columbia University; Johnston, O'Malley, Miech, Bachman, & Schulenberg, 2014). Given the predominance of adolescent substance use, informative and accurate measurement is crucial. Currently, researchers use a variety of methods to measure adolescent substance use, but these measurement techniques often produce misleading or conflicting results.

Researchers have made significant recent progress on the development of theories for the etiology and mechanisms of adolescent substance use. These theories consider a wide range of factors, ranging from biological to environmental. For example, one recent theory proposes that internalizing symptomatology during childhood and adolescence acts as a risk factor for substance use and later abuse (Hussong, et al., 2011). In this model, risk of later substance use begins in infancy, specifically with infants exhibiting behavioral inhibition and a reactive temperament. In early childhood, internalizing symptomatology is associated with less adequate interpersonal skills later in childhood. Both continued internalizing symptomatology and poor social skills are associated with peer rejection, which, taken together with adolescents' positive expectations of substances, may lead an adolescent to use substances as a coping mechanism. Onset of substance use during adolescence could then result in continued internalizing symptomatology and substance use disorders in adulthood (Hussong, et al., 2011).

Researchers have also theorized that externalizing symptomatology is a risk factor for initial substance use in adolescence and later substance abuse disorders. A large number of studies have shown that aggression and delinquent behavior are strongly associated with an earlier onset of alcohol use and with a higher number of problem behaviors later in life (Zucker, 2006). Theories of internalizing and externalizing symptomatology for adolescent substance use are not mutually exclusive. Although externalizing symptomatology has a stronger effect than internalizing symptomatology, both are important to consider and most likely have overlapping pathways (King, et al., 2004).

A variety of other pathways exist for examining adolescent substance use. Some researchers theorize that peer influence strongly affects adolescent substance use. For example, Fujimoto and Valente (2012) examined the “contagion mechanisms,” forms of communication, which influence adolescent peers to use substances. Structural equivalence (comparison of oneself to peers) rather than cohesion (direct communication) was found to more strongly influence adolescent substance use.

Recent important advances in technology allow theories of adolescent substance use to more readily incorporate biological factors. Although geneticists have made substantial progress, genetic effects are still not determined for specific substances (Kendler, et al., 2012). Even if geneticists were able to identify a substance specific gene, gene-environment interaction is significant and influences substance use (Kendler et al., 2012).

Symptomatology, environment, and biology are just several factors researchers incorporate into theories to determine the etiology and pathways of adolescent substance use. As these theories of adolescent substance use have become more complex and nuanced in recent years, there is an even greater need for a valid and reliable measure of substance use to systematically evaluate these theories. Given the wide range of current methods for measuring adolescent substance use, the results of similar studies can vary greatly depending on the particular measurement method used. Research in this area would benefit from a consistent method of measuring adolescent substance use to more accurately measure adolescent substance use and to systematically compare and contrast different theoretical models; this is my focus here.

Complexity of Dimensionality and Measurement

Dimensionality Issues.

Defining substance use is one initial issue when measuring substance use. A poly-substance user is defined as using multiple substances, although a person can no longer be diagnosed for poly-substance dependence under the DSM-V (American Psychiatric Publishing). Many researchers, however, still categorize participants as either mono- or poly-substance users despite the change in DSM criteria. One reason researchers may maintain this distinction is the underlying idea that the more substances a person uses, the more they are at risk for negative consequences, such as future addiction or mental health issues. Although a poly-substance measurement model can be quickly and efficiently measured, one weakness in this approach is that all substances are treated equally. A person who uses two legal substances (such as cigarettes and alcohol), which are common for respondents over 21, receive the same score as another person who uses two illegal substances (such as cocaine and heroin), which are likely to lead to more harmful outcomes.

The limitations of the poly-substance use model has led some researchers to use a multi-factor substance use measurement model, which allows researchers to categorize substances in a more meaningful way. These categories are potentially numerous and dependent on what the researcher chooses to examine. One possible category system is to partition substances used into legal and illegal categories. This approach has more value for substances that are legal/illegal, independent of the user's age; for example, cocaine and heroin are illegal irrespective of the user's age, whereas alcohol and cigarettes are illegal only for respondents younger than 21 (or 18). For my study, which examines adolescents (although a few participants are 18 or older), a multiple factors model based upon legality would not be sufficient because all substances are considered illegal (except for the small minority of respondents 18 years or older).

Another multi-factor substance use model would categorize substances by their biological effects, such that a researcher could separate drugs into stimulants, depressants, and others. For example, one study examined whether personality characteristics, specifically sensation-seeking behaviors and depressed mood, influenced the type of drug an adolescent used (Teichman, Barnea, & Ravav, 1989). These researchers found that drug use, regardless of biological effects, increased when adolescents endorsed sensation-seeking behaviors; whereas adolescents who exhibited anxiety depressive moods were associated with higher uses of depressants (excluding alcohol).

There is no one correct way to measure substance use, and before choosing a substance use measure (or metric), researchers need to determine the specific features of substance use that are relevant to their theoretical questions: For example, whether to treat all substances equally, focus on environmental factors such as legality, or the biological effects of substances. Different definitions will affect the measurement model a researcher uses, the ability to provide a meaningful assessment of theoretical questions, and the value of the applied implications of the research.

Measurement Issues.

Researchers have noted several limitations of current methods of the measurement of substance use, especially alcohol use. One shortcoming is the question content and structure. The reference period, or the span of time that a participant is asked to consider, usually varies between one month and 12 months for substance use items. The shorter reference period may decrease issues of memory loss and participants may report a more accurate estimate of their drinking behavior, but the 12-month period accounts for season-related drinking behavior (Dawson, 2003). Both reference periods, however, have limitations. One issue pertinent to both reference periods is that participants often estimate their drinking and the resulting data will have “random” peaks at intervals that are multiples of a probable weekly number of drinks. For example, having a peak of 20 drinks in a month may indicate that participants estimated they consumed, on average, five drinks a week and then multiplied this number by the four weeks of the month. Researches also need to consider question framing. The instructions or wording of the question may lead participants to minimize the amount of alcohol they drink to “normalize” their drinking behavior (Greenfield & Kerr, 2008).

One common way to measure alcohol use is by multiplying the average quantity of alcohol a participant drinks at a given sitting (number of drinks) by the frequency the participant drinks (QF method). The QF measurement method may be appropriate for open-ended questions, but not for the more commonly used closed questions (McGinley & Curran, 2014). The QF measurement of alcohol use may overestimate actual alcohol consumption and may change the order of relative ranks of alcohol consumption (McGinley & Curran, 2014).

Graduated frequencies (GF) are another method commonly used to measure alcohol use. The GF asks participants the number of times they have consumed a given number of standard drinks during the reference period. The measure usually begins with higher quantities of alcohol and decreases from there. The GF estimates the total amount of alcohol consumed by summing each quantity-specific frequency. Although the GF may provide a more nuanced picture of a participant’s drinking behavior, it also unintentionally allows for the possibility of calculation of a larger number of days consuming alcohol than actual number of days in the reference period (Dawson, 2003).

To a lesser extent, researchers have also examined the limitations of the measurement of illicit substance use (substances not including alcohol and nicotine). One key issue in substance use measurement is the accuracy of the report, including self-report, biological measures, and other sources of information, such as peers or parents (Carroll, 1995). The accuracy of self-report can vary greatly. Biological measures also vary in accuracy because drug use can only be detected within a certain time frame (Carroll, 1995).

There are numerous measures of adolescent substance use; for example, the Daily Drinking Questionnaire, the timeline followback, and multiplying the quantity by the frequency of a substance are all methods of measuring substance use. Moreover, many of these measures can be impractical given time constraints and lack of funds to conduct hour-long interviews with participants, thus leading researchers to use only several items on a survey or during an interview to assess adolescent substance use. This shorter method often results in a substance use score that is determined by the percentage of items a participant endorses out of the total number of substance use items. This calculation does not account for the difference in type or severity of behavior endorsed, which can present a misleading or limiting view of participants.

Researchers have proposed methods that may provide more accurate representations of actual drug use, such as measuring the actual volume and ethanol content each drink a participant consumes (Greenfield & Kerr, 2008). Although these new measurement methods may provide more accurate results, they are quite extensive and not always practical. Thus, we strive to establish a more accurate and accessible measurement model of adolescent substance use by comparing different model structures and several different scoring methods, such as classical test theory and factor analysis.

Analytic Issues.

The methods and scales used to measure substance use are diverse. A large number of the developed scales to measure drug use assess drug abuse and dependence. These measures often have certain thresholds in which a person is diagnosed with a disorder, considered at risk, or within a normal range. These measures are sometimes less concerned with the precise frequency of drug use or which combination of drugs are used, but more interested in having sufficient information to place someone within a certain category. These assessment tools are needed for clinical diagnoses and some areas of research, but are not necessarily sufficient for all research of substance use.

National organizations recommend certain scales because of the wide range of scales available so “researchers can more easily compare and combine datasets to detect more subtle and complex associations among variables, thereby promoting greater collaboration, efficiency, and return on investment” (National Institute on Drug Abuse, 2014). For example, The National Institute on Alcohol Abuse and Alcoholism (NIAAA) recommends several scales to measure adolescent drug use, such as the Adolescent Alcohol Involvement Scale (AAIS), the Alcohol Timeline Followback (TLFB), the Drug Use Screening Inventory (DUSI-R), and the Substance Abuse Subtle Screening Inventory (SASSI) (Winters, 2003). The NIAAA also recommends a set of three to six questions to assess a person’s patterns of alcohol consumption. These three to six questions ask about frequency of alcohol use, maximum drinks consumed at one time period, frequency of binge drinking, and the typical number of drinks consumed on a typical day when drinking alcohol (National Institute of Alcohol Use and Alcoholism, 2003). A majority of these items are binned counts (e.g., 0, 1-3, 4-6, etc.), which, as mentioned previously, are problematic (McGinley & Curran, 2014). These items were also established over 10 years ago, and given the new developments and complexities of theories of adolescent drug use, new guidelines to assess alcohol use and other drugs may be necessary.

Scoring methods in current adolescent drug use research often deviate from the numerous recommended scales. Researchers use a wide range of methods to collect data on alcohol and drug use, such as raw count of the number of drug used, calculating the quantity and frequency of the drug use, or assessing whether a person is a heavy drinker. For example, Haller et al. (2010) examined the relationships between adolescent substance use, association with peers who endorse substance use, and low levels of academic achievement over an 18-year time period on adults with substance use disorders. Within their study, these researchers used multiple measures to assess drug and substance use disorders as set by the DSM-III-R. Participants reported their frequency of binge drinking on an eight-point scale from zero (never) to seven (every day) to assess binge drinking (consuming five or more drinks at a single sitting) during three of the time points. Participants also reported the frequency of use of eight different classes of drugs to assess drug use, also at the same three time points as the binge drinking items. Haller and colleagues used a composite poly-substance use method to assess drug use, which given the goal of the study seems appropriate, but also has limitations.

Abar et al. (2014) also examined substance use, specifically alcohol use, at a transition point between adolescence and adulthood and studied whether the parent-child relationship for students in their first year of college affected the alcohol-related behaviors these students exhibited during their first year of college. These researchers assessed the parent-child relationship with several measures, but also measured alcohol consumption using the Daily

Drinking Questionnaire (DDQ) and with the Heavy Episodic Drinking (HED) measure as recommended by the NIAAA. The DDQ asks participants to report the average number of drinks they consume on a typical Friday or Saturday night over the past months; these drinks are then summed to determine a total score for the participant. The HED asks participants to report how often they participated in binge drinking over the past two weeks. Unlike the scoring methods used in the study above, this study distinguishes not only frequency of drinking, but also more closely examines heavy drinking behaviors.

Haller et al. (2010) used a count method of scoring adolescent substance use that treats each substance used equally, approaching substance use within the poly-substance use framework. Similar to classical test theory in which each item is treated equally, each substance is scored dichotomously and the scores from each item are summed. The resulting score for each participant is considered the observed score, which is a combination of the true underlying score and error (Hambleton & Jones, 1993). Although scoring methods based on classical test theory have its advantages, such as weak assumptions to test the model, they also have several weaknesses, such as the lack of discrimination of difficulty (or differentiability of item weights) (Hambleton & Jones, 1993).

In contrast, factor analysis accounts for the differential contributions of each item to an underlying construct. This method allows researchers to approach the measurement of substance use by allowing for different weights for each substance. More specifically, in this case, researchers would use non-linear factor analysis, which creates scores for each participant while accounting for the differential strength between each item and the underlying construct (Curran et al., 2014). Key advantages to non-linear factor analysis include the ability to account for non-linear relationships between the items and the underlying latent factor, the assumption that the latent factor is the primary origin of all of the items, and the establishment of a standard normal scale for the resulting factor scores (Hussong, Curran, & Bauer, 2013). If a researcher were to only examine the binary use or non-use of a substance (not including the frequency), then this factor analysis would be equivalent to examining the items using item response theory (IRT).

As these examples illustrate, there are a wide variety of methods to measure adolescent substance use. The measures within these examples reflect three basic categories: count of substances endorsed, division of users based upon relative amount of substance use (users versus people who binge use), and division of users based on a quality, such as the biological effects, of the substance (Haller et al., 2010; Abar et al., 2014; Teichman, Barnea, & Ravav, 1989). In the current study, I examined three scoring methods for adolescent substance use similar to methods used in previous studies: a measure that dichotomously distinguishes substance users from substance abstainers, a proportion of the number of different substances endorsed, and factor scores based on the specific pattern of individual substances endorsed.

Hypotheses

I used assessments of lifetime use (users versus non-users), the proportion of substances endorsed, and factor scores as predictors of academic achievement and internalizing and externalizing symptomatology using a previously established data set, The National Longitudinal Study of Adolescent to Adult Health (Add Health[1]). I examined data collected during the first

---

[1]This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

wave of Add Health to compute the scores for each measurement model and then compared these three different scoring methods based on their differential predictability for the three aforementioned outcome variables.

First, I compared the distributions of scores for each measurement model. Specifically, I examined the differences between the proportion model (the proportion of substances a participant used) to the factor score model (the specific combination of substances a participant used). Finally, I used these three different models to compare and contrast the predictive accuracy of each model as determinants of academic performance and internalizing and externalizing symptomatology. I explored whether race, gender, and age differences affect the substance use measurement models.

I predicted that the proportion model would account for a larger portion of the variance within each model, such that, holding race, gender, and age constant, the proportion model would be a better predictor of academic achievement and internalizing and externalizing symptomatology as compared to the lifetime use model, which only distinguishes between substance users and non-users. Further, I anticipated the factor score model would better account for the differences in academic achievement and internalizing and externalizing symptomatology above and beyond those predicted by the proportion model. I posited that the factor analytic approach would be a more accurate model because it would account for the differences in substances by differentially weighting each item. The factor analytic approach would offer a more refined view of substance use by considering the type of substance endorsed.

Method

Participants

My study used the public-use data set available from The National Longitudinal Study of Adolescent to Adult Health (Add Health). The researchers of Add Health selected 80 high schools and 52 middle schools to create a representative sample based on region and population density of the United States, school size and type, and ethnicity (Harris et al., 2009). The participants used here are a sub-sample of the entire Add Health data set (N= 20,745) for “In-home” interviews at Wave I (Harris et al., 2009). This public-use sub-sample contains 6,504 participants, however, I deleted 647 respondents who did not provide valid responses to all items. My final sample consisted of 5,857 participants. I used list-wise deletion to determine my final sample; although list-wise can create biases, my focus is on contrasting specific measurement models rather than generalizing to the population. The final data set, which contains no missing data, will allow for a more accurate conclusion that the independent variable (the method of substance use measurement) affects the final outcome. The interviews for Wave I were conducted in 1994 and 1995, and the participants ranged in age from approximately 11 to 21 years old (born between the years 1974 and 1983), with an average age of 16. At Wave I, approximately 48% of the participants are male and a majority of the participants identify as white (67%). See Table 1 for descriptive statistics on all primary variables within the study.

Measures

I used specific sections of the Add Health data set including basic demographic questions, substance use items, academic achievement (grades in specific classes), assessment of internalizing symptomatology (through the feelings scale), and assessment of externalizing symptomatology (through the delinquency items). See the appendix for the specific items used.

Substance Use.

The section of primary interest focused on substance use and contained a total of 54 items; however, I selected seven items that pertain to the lifetime use of seven different classes of drugs: cigarettes, chewing tobacco, alcohol, marijuana, cocaine, inhalants, and other illicit drugs. I chose not to examine the remaining items because these items did not directly pertain to lifetime substance use. For example, the excluded items assessed recent substance use, peer substance use, and the availability of substances. The response scales for the seven items are dichotomous, categorical, or ordinal. I used these seven items to create substance use measurement models using three different methods. These seven items ask about lifetime use of each drug, which includes the substances: cigarettes, chewing tobacco, alcohol, marijuana, cocaine, inhalants, and other illicit drugs. One example of an item is on lifetime use of alcohol, which is assessed through the question "Have you had a drink of beer, wine, or liquor – not just a sip or a taste of someone else's drink – more than two or three times in your life?" The responses of this item are dichotomous: yes or no. All seven items allowed participants to answer "don't know" and "not applicable," and participants were permitted to refuse to answer a question. If any participant selected any of the three previous responses, that participant was removed from the data set. All seven items were dichotomized, such that each item was scored as either a zero or one: a participant either never used a specific drug in their lifetime or the participant used a specific drug in their lifetime. The seven items were then scored according to the specific model: The lifetime use model grouped participants into substance abstainers and substance users, the proportion model calculated a proportion of the seven substances participants used, and the factor score model used the seven items in a confirmatory factor analysis to create the scores.

Measurement Model 1: Lifetime Use.

The first measurement model distinguishes lifetime substance abstainers from lifetime substance users. The seven dichotomized substance use items were used to determine whether participants had ever used a substance, and thus separated participants into one of two groups. Participants who used at least one substance (number of substances used in their lifetime was greater than or equal to one) were labeled as substance users, whereas participants who had never used any substance (number of substances used in their lifetime was equal to zero) were labeled as substance abstainers. Approximately 69% of participants reported having used at least one substance in their lifetime.

Measurement Model 2: Proportion.

The second measurement model is an unweighted proportion of lifetime substance use. This measurement model used the same seven binary items that assessed lifetime substance use. Like the lifetime use model, all items were dichotomized. The response from each item was then summed and divided by seven to determine a final lifetime drug use measure. The total lifetime scores range between zero and one, which represents the number of substances used. The average number of substances used during a participant's lifetime is 0.22 (M = 0.22, SD = 0.20), which is between one and two different substances.

Measurement Model 3: Factor Scores.

The third method of measurement used non-linear factor analysis. This measurement model used the same seven substance use items that were used in the lifetime use and proportion measurement model. Using Mplus, these seven binary items were used in a non-linear confirmatory factor analysis to calculate factor scores using maximum likelihood estimation. The obtained factor score estimates are associated with one factor, the underlying latent variable that

I label as poly-substance use. Factor scores were standardized (*M* = 0, *SD* = 0.99) and ranged between -1.17 and 3.01.

Academic Achievement.

Academic achievement is measured by 24 items that include questions on skipping a grade, grades in certain classes, expulsion, and interactions with teachers and peers at school. I selected only four of these items to assess academic achievement: Grades in English, math, social studies, and science classes. One example of these items is “At the most recent grading period/last grading period in the spring, what was your grade in English or language arts?” The responses for this item include A, B, C, D or lower. These four items also offered the responses “didn’t take this subject,” “took subject/wasn’t graded this way,” and “don’t know,” and participants were permitted to refuse to answer the questions, all of which were not included in the calculation of GPA. Although I used list-wise deletion to determine my final sample for all other items within the study, I did not use list-wise deletion to determine grade point average. Instead, I calculated GPA for any participant who had at least one class. If a participant did not have any valid answers for at least one class, they were removed from the data set. The average GPA is approximately 2.82 with a standard deviation of 0.76 (*M* = 2.82, *SD* = 0.76) and range between 1 and 4.

Internalizing Symptomatology.

Internalizing symptomatology is assessed with 19 items drawn from the “Feelings Scale” within Add Health. The feelings scale contains questions about subjective feelings, such as happiness, depression, loneliness, and enjoyment of life on a four-point scale of “never/rarely” to “most/all the time,” with the options of “refused” and “don’t know.” Items include “You were bothered by things that usually don’t bother you,” “you felt depressed,” and “you felt hopeful about the future." This scale is essentially the Center for Epidemiologic Studies Depression Scale – Revised (CESD-R) minus one item on sleep quality. The CESD-R assigns a score ranging from 0 to 3 for each item, with reverse scoring for appropriate items. This scale has certain thresholds for diagnoses, but in general, the higher a participant's score, the more severe the internalizing symptomatology (Eaton, et al., 2004). I summed the score from each item and then averaged to calculate a final score for internalizing symptomatology for each participant, with total scores ranging from zero (exhibiting no signs of internalizing symptomatology) to 3 (endorsing all items at the highest level for internalizing symptomatology). The average internalizing symptomatology score is 0.56 (on a scale ranging from zero to three) with a standard deviation of 0.39 ($M = 0.56$, $SD = 0.39$) and range between zero and 2.63.

Externalizing Symptomatology.

Externalizing symptomatology was assessed using the 15-item delinquency scale within the Add Health data set. These items ask about stealing, selling drugs, participating in physical fights, and carrying a weapon on a four-point scale of "never" to "5 or more times." Several examples of items include "In the past 12 months, how often did you deliberately damage property that didn't belong to you?", "How often did you get into a serious fight?", and "In the past 12 months, how often did you steal something worth more than $50?" The responses from each item were summed and then averaged (*never* = 0 to *5 or more times* = 3) to determine a total score for delinquency. Scores ranged from zero to three, with low scores indicating less delinquent behaviors. The average delinquency score is approximately 0.26 (on a scale ranging from zero to three) with a standard deviation of 0.31 ($M = 0.26$, $SD = 0.31$) and range between zero and 2.80.

Race.

The Add Health study collected extensive data on race. For my study, however, I included race as a binary variable: a participant is either a member of the racial majority or a member of the racial minority. I recognize that race can play a significant and important role within substance use (both frequency of use and type of substance); however, race is not a key focus of my study. I still include a more limited race variable to acknowledge possible racial differences in substance use, but I believe future research should further examine the differences between race and substance use. Approximately 32.8% of the sample is a racial minority (Table 1).

Age.

Add Health respondents ranged between 12 and 21 years old. The average age for the study is 15.98 with a standard deviation of 1.75 ($M = 15.98$, $SD = 1.75$): see Table 1. The availability of substances is likely to vary by age. To reduce this effect, I include age as a covariate within the regression models.

Gender.

Approximately 48% of the sample is male. Prior research (Byrnes, et al. 1999) suggests there are gender differences in risk-taking, which is often related to substance use, thus I included gender within the regression models to control for this effect.

Plan of Analyses

I began by creating the substance use scores for each substance use measurement model. I first compared the distributions of the proportion and factor score models, looking specifically at the differences in substances used per proportion score. I then used each substance use measurement model (lifetime use, proportion, or factor scores) as a predictor variable, along with age, race, and gender, in a multiple linear regression model. I first examined the contributions of age, race, and gender on academic achievement and internalizing and externalizing symptomatology using multiple linear regressions. I then used these factors as covariates and added one of the three substance use measurement models to predict academic achievement and internalizing and externalizing symptomatology, which resulted in nine unique regression models. Finally, I examined the squared semi-partial correlations of each predictor variable within each regression model.

Results

Measurement Model Comparisons

Univariate Distributions.

The first measurement model, lifetime use, determined whether participants were substance users or substance abstainers. The average score for lifetime use was 0.69, indicating that 69% of participants had tried at least one of the seven drugs included in this study (M = .69, SD = .46).

The second measurement model, the proportion model, not only distinguished between substance abstainers and substance users, but also accounted for the number of drugs a participant used in their lifetime. The average proportion of substances used was 0.22, which indicates that, on average, participants endorsed trying between one and two of the seven items assessed in this study (M = .22, SD = .20). The distribution is positively skewed, such that a larger number of participants endorsed using a smaller proportion of substances. The distribution of the scores from this model can be seen in Figure 1.

The third measurement model, the factor score model, further distinguished substance users by differentially weighting each substance to determine each participant's final factor score. Similar to the proportion model, the distribution of factor scores is positively skewed (although there is a slight uptick of scores at the high end), indicating that a larger portion of participants had lower factor scores. A depiction of the distribution of factor scores can be seen in Figure 2. Factor scores varied on the combination of substances used. Overall, factor scores ranged from -1.17 to 3.01 ($M = 0$, $SD = 1$). For example, if a participant endorsed using two of the seven substances, then the factor scores could be 0.35 if the two substances were cigarettes and alcohol, 0.61 if the two substances were alcohol and marijuana, or 1.05 if the two substances were alcohol and illicit drugs. The distribution of the factor scores for participants who selected two substances can be seen in Figure 3. As can be seen within the figure, the distribution of factor scores does not align with the distribution of the proportion model. Figures 1, 2, and 3 help illustrate that the proportion and factor score model are not equivalent. A comparison of Figures 1 and 2 show that the overall distribution of scores, while both positively skewed, do differ in that the factor score distribution does not taper off like the proportion model. Figure 3 depicts the factor score distribution for a single proportion score, showing that the factor score model provides greater variability and a more refined scale than the proportion model. The differing distributions highlight that not only the number of substances endorsed, but also *which* substances were endorsed matters.

Bivariate Distributions and Correlations.

The relationships between the lifetime use model, the proportion model, and the factor score model are positively correlated (see Table 2 for more detail on the correlations between all primary variables within the study). The relationship between the lifetime use model and both the proportion and factor score model are not particularly telling, given that the lifetime use model only has two categories. The relationship between the proportion and factor score models, however, is more informative. The two models are positively correlated and the variance between the two models decreases with higher scores (Figure 4).

As expected, the measurement models within this study are highly correlated, which creates similar abilities of each measurement model to predict certain outcomes as can be seen in the regression analyses below. The measurement of lifetime substance use was significantly correlated with both the proportion model ($r = .72$, $p < .0001$) and the factor score model ($r = .78$, $p < .0001$). The proportion model and the factor score model were also highly correlated ($r = .98$, $p < .0001$). It is worth noting, however, that although the three measurement models are highly correlated, correlations take into account the rank-order of each participant as compared to others, but does not take into account the variability amongst participants. Further examination is needed to fully understand similarities and differences among these scores.

Regression Analyses

Covariate Regressions.

For each regression model, I included age, gender, and race as covariates to separately predict internalizing and externalizing symptomatology and academic achievement. I estimated three regressions first using just the three covariates to examine the effects of these variables prior to assessing the substance use measure within each model.

Age, race, and gender significantly accounted for variation within grade-point average ($R^2 = .041$, $F(3, 5853) = 84.02$, $p < .0001$). Age, race, and gender were all significant, unique predictors of academic achievement (gender: $\beta = -.23$, $t(5853) = -12.23$, $p < .0001$; race: $\beta = -.16$, $t(5853) = -7.67$, $p < .0001$; age: $\beta = .35$, $t(5853) = 6.22$, $p < .0001$): see Table 3. More specifically, males and members of the ethnic minority were more likely to have a lower grade-point average, whereas grade-point average increased with age.

The three covariates, age, race, and gender also accounted for a significant amount of the variance within internalizing symptomatology ($R^2 = .04$, $F(3, 5853) = 79.98$, $p < .0001$). All three covariates were again significant, unique predictors of internalizing symptomatology (gender: $\beta = -.09$, $t(5853) = -9.38$, $p < .0001$; race: $\beta = .09$, $t(5853) = 8.32$, $p < .0001$; age: $\beta = -.03$, $t(5853) = -9.24$, $p < .0001$), such that males were more likely to report lower levels of internalizing symptomatology than females, ethnic minorities were more likely to have higher levels of internalizing symptomatology than those of the ethnic majority, and levels of internalizing symptomatology decreased with age (Table 4).

Finally, age, race, and gender significantly accounted for variance within externalizing symptomatology ($R^2 = .02$, $F(3, 5853) = 40.62$, $p < .0001$). These three covariates significantly predicted externalizing symptomatology (gender: $\beta = .09$, $t(5853) = 10.82$, $p < .0001$; race: $\beta = .02$, $t(5853) = 2.12$, $p < .05$; age: $\beta = 0$, $t(5853) = -.48$, $p = .63$), indicating that males and those of the ethnic minority were more likely to participate in delinquent behavior than females; however, age was not a significant predictor of delinquency (Table 5).

As seen in these three regressions, age, race, and gender, in most cases, contribute to academic achievement and internalizing and externalizing symptomatology. Although these factors influence these outcomes, and will be retained in my subsequent model tests, my focus next is on the three approaches to the measurement of substance use. I will more closely examine the role of these measurement models in predicting the outcome measures.

Academic Achievement.

I estimated three multiple linear regressions predicting academic achievement, using gender, race, and age and a distinct measurement of substance use as predictors. The first regression used the lifetime use measurement model of substance use. The regression model explained 7.72% of the variance in grade-point average (R^2 = .077, $F(4, 5852) = 122.54$, $p <$ .0001). The lifetime use model significantly predicted grade-point average (unstandardized-β = -.32, $t(5851) = -15.11$, $p < .0001$), indicating that lifetime substance users had significantly lower GPAs than substance abstainers (Table 3). The squared semi-partial correlation of lifetime substance use was .036, indicating the substance use measure uniquely accounted for approximately 3.6% of the variance of GPA.

The second regression included the proportion measurement model as a predictor. The regression model explained 9.13% of the variance in GPA (R^2 = .091, $F(4, 5852) = 147.08$, $p <$ .0001). The proportion model predicted GPA (β = -.87, $t(5851) = -17.96$, $p < .0001$), such that GPA decreased as the number of substances used increased (Table 3). The squared semi-partial correlation of the number of substances used is .05, which is higher than the squared semi-partial of the lifetime use model (5% versus 3.6%).

The final regression used the factor scores of each item on the poly-substance use latent factor as a predictor of GPA. The regression model explained 9.45% of the variance within GPA (R^2 = .094, $F(4, 5852) = 152.71$, $p < .0001$).T he factor score model significantly predicted GPA (β = -.18, $t(5851) = -18.55$, $p < .0001$), such that GPA decreased as factor scores increased (Table 3). The squared semi-partial correlation of substance use within the factor score model is equal to .053, which is only slightly larger than that obtained using the proportion score.

As expected, all three measures of substance use significantly predicted GPA. Consistent with expectations, both the proportion model and the factor score model (with the factor score model being the more predictive of the two) accounted for more of the variance within the academic achievement model than the lifetime use model of substance use.

Internalizing Symptomatology.

I estimated three multiple linear regressions predicting internalizing symptomatology. Each regression included gender, race, and age and one of the three measures of substance use. The first regression used the lifetime use model of substance use as a predictor. The regression model explained approximately 6.84% of the variance within internalizing symptomatology ($R^2 = .068$, $F(4, 5852) = 107.57$, $p < .0001$). The lifetime use model significantly predicted internalizing symptomatology ($\beta = .149$, $t(5851) = 13.52$, $p < .0001$), indicating that substance users had higher levels of internalizing symptomatology than substance abstainers, with a squared semi-partial correlation of lifetime substance use equal to .029 (Table 4).

The second regression used the proportion model of lifetime substance use as a predictor. The regression model explained 9.27% of the variance within internalizing symptomatology ($R^2 = .092$, $F(4, 5852) = 149.49$, $p < .0001$). The proportion model significantly predicted internalizing symptomatology ($\beta = .461$, $t(5851) = 18.55$, $p < .0001$), such that levels of internalizing symptomatology increased as the proportion of substances used increased, with a squared semi-partial correlation of the number of substances used of .053 (Table 4).

The third regression used the factor scores of each item on the poly-substance use latent factor as a predictor of internalizing symptomatology. The regression model explained 9.4% of the variance within internalizing symptomatology ($R^2 = .094$, $F(4, 5852) = 151.80$, $p < .0001$), with a squared semi-partial correlation of poly-substance use factor scores equal to .05. The factor score model significantly predicted internalizing symptomatology ($\beta = .096$, $t(5851) = 18.78$, $p < .0001$), such that as factor scores increased, so did levels of internalizing symptomatology (Table 4).

Similar to academic achievement, all three measurement models of substance use significantly predicted internalizing symptomatology. The proportion and factor score models, however, predicted more of the variance than the lifetime measure of substance use, as indicated by the semi-partial correlations and the overall improvement of model fit (with the factor score model being the more predictive of the two).

Externalizing Symptomatology.

Finally, I estimated three multiple linear regressions predicting delinquency. Each regression included gender, race, age, and one of the three measurement models of substance use. The first regression used the lifetime use model of substance use as a predictor. The regression model explained 12.13% of the variance within delinquency (R^2 = .121, F(4, 5852) = 202.04, p < .0001). The lifetime use model significantly predicted delinquency, (β = .23, t(5851) = 25.93, p < .0001), indicating that substance users were more likely to commit delinquent acts than abstainers, with a squared semi-partial correlation of lifetime substance use of .10 (Table 5).

The second regression used the proportion model of substance use as a predictor. The regression model explained approximately 25.47% of the variance within delinquency (R^2 = .254, F(4, 5852) = 499.96, p < .0001). The proportion model significantly predicted delinquency (β = .79, t(5851) = 42.89, p < .0001), such that participants who used a larger proportion of substances were more likely to commit delinquent acts, with a squared semi-partial correlation of the proportion of substances used of .234 (Table 5).

The third regression used the factor scores of each item on the poly-substance use factor as a predictor of delinquency. The regression model explained 25.29% of the variance within delinquency (R^2 = .252, F(4, 5852) = 495.41, p < .0001), with a squared semi-partial correlation of poly-substance use factor scores of .232. The factor score model significantly predicted delinquency ($\beta = .16$, $t(5851) = 42.68$, $p < .0001$), such that participants with higher factor scores were more likely to commit a larger number of delinquent acts (Table 5).

All three measures of substance use were significant predictors of delinquency. When the proportion model and the factor scores model were used, the overall model fit increased; the predictive accuracy of substance use within each model also increased when using the proportion and factor score models. The measure of lifetime use predicted less then half of the variance than the proportion and factor score models, accounting for only 10% of the variance as opposed to approximately 23% of the variance, as indicated by the squared semi-partial correlations. Unlike the previous two outcome measures (academic achievement and internalizing symptomatology), the proportion model account for slightly more of the variance in the overall model than the variance account by the factor score model based upon the squared semi-partial correlations (23.4% versus 23.2%), although this difference is negligible.

Discussion

Adolescent substance use is a prevalent issue that has numerous adverse consequences (The National Center on Addiction Substance Abuse at Columbia University; Johnston, O'Malley, Miech, Bachman, & Schulenberg, 2014). Given its pervasiveness and connection to a wide range of negative outcomes, many theories incorporate adolescent substance use within their models. The empirical measurement of substance use within these models, however, varies. Researchers examine lifetime and recent substance use, the frequency of use, the number of substances used, and the type of substances used, and combine one or several of these items in different ways. The field of psychology lacks a consistent and valid measure of adolescent substance use that could be used in the growing number and complexity of current theory.

Although there are many approaches to the measurement of adolescent substance use, I focused on the assessment of lifetime adolescent substance use. More specifically, I examined the differential predictability of three measurement models of substance use on academic achievement and internalizing and externalizing symptomatology. These models differed in the level of specificity with which they distinguished substance users. The first model separated substance users from substance abstainers. The second model distinguished substance users by the number of substances used. Finally, the third model distinguished the substance users by the number and *types* of substances used.

Measurement Models

My first analyses examined the underlying distributions of the three measures of substance use. The lifetime use model separated participants into users and non-users. This resulted in a dichotomous measure with approximately two-thirds of participants having used at least one substance. The proportion model distinguished substance users by the number of substances endorsed. These scores resulted in a positively skewed distribution, indicating a majority of participants used two or fewer substances. The factor score model distinguished participants by the type of substances used. Similar to the proportion model, the distribution of factor scores was positively skewed.

Although the proportion and factor score models have similar distributions, these distributions represent different information. The proportion model represents the number of substances endorsed. Each item is weighted equally; a one if endorsed and a zero if not endorsed. Participants who had a higher score used more substances than a participant with a lower score. The scores of the factor score model, however, are not on the same scale and are not directly comparable to the scores of the proportion model. These scores provide a more refined estimate of poly-substance use by weighting the response to each item by the factor loading for that item.

The distributions of the proportion model and the factor score model were positively skewed, but, as mentioned previously, these models represent the measurement of poly-substance use differently for each respondent. For example, when applying the proportion model for participants who used two substances, the proportion score is 0.29; thus, the participants who endorsed two substances will receive the same score regardless of the type of substance used. When looking at the factor scores from these same participants, however, the distribution is more varied. These differences suggest the type and combination of substances does matter: different combinations result in a wide range of factor scores, which is information not captured by the proportion model.

The varying distributions among the measurement models show each model represents different aspects of the latent variable of substance use. The lifetime use model is a relatively unrefined measure, whereas the proportion and factor score models are more nuanced. The proportion and the factor score models, however, depict different facets of substance use. More specifically, these distributions show that the number and the type of substances endorsed do not represent the construct in the same way. Overall, these initial results help to illustrate the potential differences in the predictive accuracy of each substance use measurement model.

Regressions

My primary analyses focused on multiple linear regressions in which one of the three measurement models (lifetime use, proportion, and factor score) was used as an independent variable (along with age, race, and gender) to predict academic achievement or internalizing or externalizing symptomatology. I estimated an initial regression in which age, race, and gender were the only predictors. As expected, all three were significant predictors of academic achievement and internalizing and externalizing symptomatology (although age was not a predictor of externalizing symptomatology).

I then extended these three regressions to include substance use for each of the three outcomes (academic achievement and internalizing and externalizing symptomatology). As expected, for all three outcomes, the lifetime use measurement model was less precise than the proportion and factor score models. For two of the three outcomes, the factor score model improved predictability of the outcome measure over the proportion model. The differences between the proportion model and factor score model, however, were often small. For both academic achievement and internalizing symptomatology, the factor score model predicted slightly more of the variance within the model, whereas, for externalizing symptomatology, the proportion model predicted slightly more of the variance within the model.

The results of these regression analyses are consistent with current theory. In previous research, substance use was associated with decreased academic achievement and increased levels of internalizing and externalizing symptomatology. Although all three measurement models of substance use were significant predictors of academic achievement and internalizing and externalizing symptomatology, the measurement of substance use affected the precision of the regression model. The proportion and factor score models were notably more accurate than the lifetime use model. Given these differences, the proportion and factor score models should be implemented instead of the lifetime use model. In practice, psychological theory is becoming more complex and nuanced and sample sizes are likely to be much smaller than this particular study, thus researchers need a more precise measure that will detect the effect of substance use.

Potential Limitations and Future Directions

There are several limitations in this study. First, the subset of substance use items selected for this study was limited by the items included in the Add Health study, such as lifetime and recent substance use, binge-drinking, peer substance use, and accessibility of certain substances. I chose to focus on lifetime substance use because base rates of less frequently used drugs would be higher, and availability is an influential factor of adolescent use. Adolescents have more limited access to all of these substances (if they are under the age of 18) and may only use them when they are available, which may not be in the range covered by substance use items that assess recent behaviors. Second, the items only assess seven substances on a self-report measure. Items that cover a wider range of substances, rather than an "illicit drugs" category, would be more informative. Self-report measures also have many limitations; for example, participants may want to portray themselves in a more favorable light and not report using certain substances.

Third, the manner in which I treated the covariates within the regression models (age, race, and gender) limited the conclusions. I dichotomized race by grouping participants as either part of the racial majority or racial minority, which limited conclusions based on race. A wide range of other important factors could have also been included in the regression analyses, such as socioeconomic status, family history of substance use, and peer substance use. Finally, the data I examined were cross-sectional in nature and I cannot conclude causality between substance use and any of the three outcome variables. To do so, I would need to examine changes in substance use and the three outcome variables for the individual over time. For example, substance use may cause a participant's grade-point average to decrease; however, a participant's grades could be declining, which then leads the participant to use substances.

The measurement of adolescent substance use has much room for future improvement. Alternative methods for determining poly-substance use should also be examined more closely. For example, measurement models that also include recent use and frequency of use may be more informative than models that only include lifetime use. The measure of lifetime substance use may be less predictive of certain outcomes over time, given the number or types of substances endorsed may not change over time, but frequency or amount of use may change. Taken together, the measurement of adolescent substance use needs to be studied in greater depth and in a variety of different ways.

Recommendations for Applied Researchers

Given the results of the regressions, I would recommend using the proportion model or factor score model rather than the lifetime use model. The proportion and factor score model more accurately measure poly-substance use, the intended construct, by using a more nuanced assessment of substance users. The proportion and factor score model had higher levels of predictability for all three outcomes within the regression models. A recommendation of the factor score model over the proportion model, however, is less clear. Although, the factor score model had the highest predictability for two of the three outcomes, the added complexity (and associated assumptions) of the factor analytic approach may limit its value over the slight improvement over the proportion model. Factor analysis, in general, is a more demanding statistical procedure that requires more advanced statistical knowledge. The results of factor weights may also be difficult to interpret beyond a general conclusion of one substance having a greater weight than another. Finally, factor analysis does not account for individual differences in the weight of each factor score. Given these more stringent requirements and several other limitations, the minor improvement of the factor score model over the proportion model may not warrant its use.

The proportion model, similar to the factor score model, also has its own advantages and disadvantages. The scores within the proportion model can be calculated more easily. The proportion model, however, does not differentially weight the impact of each drug. I do not expect all researchers to view the use of certain drug similarly. For example, adolescents drinking alcohol would be more socially acceptable or commonplace than adolescents using cocaine. Thus, the differences among the social and biological effect of these drugs suggest *not* treating them as equivalent.

Given the advantages and disadvantages of the proportion model and factor score model, I recommend that the choice of measurement model should ultimately rest on theory. If a theory incorporates the social or biological differences associated with particular substances, then factor scores should be used. If a theory seeks to examine overall drug use regardless of the type of drug, then the proportion model should be used. Although the proportion model and factor score model obtain similar results in this instance, it may not always be the case, thus measurement should be consistent with theory.

Conclusion

Taken together, the measurement of adolescent substance use is complex, dependent on theory, and can be determined in a variety of methods. My study empirically tested the differences of three measurement models of substance use on outcomes that are typically associated with substance use in current psychological theory, including academic achievement and internalizing and externalizing symptomatology, while holding all other variables (substance use items, covariates, the sample) constant. The three measurement models were relevant to current theory and were also calculated in ways that current research often examines substance use (by number of substances endorsed and the type of substances used). My study demonstrated the importance of distinguishing substance users by either the number of substances used or by the types of substances used. Although the differential predictability of these models is similar, the method of measurement should be based on relevant theory. In sum, measurement of substance use matters when predicting academic achievement and health outcomes.

References

Abar, C. C., Turrisi, R. J., & Mallett, K. A. (2014). Differential trajectories of alcohol-related behaviors across the first year of college by parenting profiles. *Psychology Of Addictive Behaviors*, *28*(1), 53-61. doi:10.1037/a0032731

American Psychiatric Association, American Psychiatric Publishing. (2013). *Highlights of changes from DSM-IV-TR to DSM-5*. Retrieved from http://www.dsm5.org/Documents/changes%20from%20dsm-iv-tr%20to%20dsm-5.pdf

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin,125*(3), 367-383. Retrieved from http://search.proquest.com/docview/57376727?accountid=14244

Carroll, K. M. (1995). Methodological issues and problems in the assessment of substance use. *Psychological Assessment, 7*(3), 349-358. doi:10.1037/1040-3590.7.3.349

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., Sher, K., Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate Behavioral Research, 49*(3), 214-231. doi: 10.1080/00273171.2014.889594

Deborah A. Dawson. (2003). Methodological issues in measuring alcohol use. *Alcohol Research and Health, 27*(1), 18-29. Retrieved from http://search.proquest.com/docview/222390247?accountid=14244

Eaton, W.W., Muntaner, Smith, C., Tien, A., Ybarra, M. (2004). Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-r). In Maruish, M.E. (ed.), *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment* (pp. 363-377). Mahwah, NJ: Larence Erlbaum.

Fujimoto, K., & Valente, T. W. (2012). Social network influences on adolescent substance use: Disentangling structural equivalence from cohesion. *Social Science & Medicine, 74*(12), 1952-1960. doi:10.1016/j.socscimed.2012.02.009

Greenfield, T. K., & Kerr, W. C. (2008). Alcohol measurement methodology in epidemiology: Recent advances and opportunities. *Addiction, 103*(7), 1082-1099. doi:http://dx.doi.org.libproxy.lib.unc.edu/10.1111/j.1360-0443.2008.02197.x

Haller, M., Handley, E., Chassin, L.& Bountress, K. (2010). Developmental cascades: Linking adolescent substance use, affiliation with substance use promoting peers, and academic achievement to adult substance use disorders. *Development and Psychopathology, 22*(4), 899-916. doi:10.1017/S0954579410000532.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47. doi:10.1111/j.1745-3992.1993.tb00543.x

Harris, K.M., C.T. Halpern, E. Whitsel, J. Hussey, J. Tabor, P. Entzel, & Udry, J.R. 2009. The National Longitudinal Study of Adolescent to Adult Health: Research Design [WWW document]. URL: http://www.cpc.unc.edu/projects/addhealth/design.

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology, 9*, 61-89. doi:10.1146/annurev-clinpsy-050212-185522

Hussong, A. M., Jones, D. J., Stein, G. L., Baucom, D. H., & Boeding, S. (2011). An internalizing pathway to alcohol and substance use disorders. *Psychology of Addictive Behaviors, 25*(3), 390-404. doi:10.1037/a0024519

Kendler, K. S., Chen, X., Dick, D., Maes, H., Gillespie, N., Neale, M. C., & Riley, B. (2012). Recent advances in the genetic epidemiology and molecular genetics of substance use disorder. *Nature Neuroscience, 15*(2), 181-189. doi:10.1038/nn.3018

King S.M., Iacono W.G., McGue M. (2004). Childhood externalizing and internalizing psychopathology in the prediction of early substance use. *Addiction*. *99(12)*:1548–1559. Retrieved from http://search.proquest.com/docview/57182249?accountid=14244

Johnston, L.D., O'Malley, P.M., Miech, R.A., Bachman, J.G., & Schulenberg, J.E. (2014). *Monitoring the future national results on drug use: 1975-2013: Overview, Key Findings on Adolescent Drug Use.* Ann Arbor: Institute for Social Research, The University of Michigan.

Leccese, M., Waldron, H.B. (1994). Assessing adolescent substance use: A critique of current measurement instruments. *Journal of Substance Abuse Treatment, 11*(6), 553-563. Retrieved from http://search.proquest.com/docview/61622107?accountid=14244

McGinley, J. S., & Curran, P. J. (2014). Validity concerns with multiplying ordinal items defined by binned counts: An application to a quantity-frequency measure of alcohol use. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 10*(3), 108-116. doi:10.1027/1614-2241/a000081

The National Center on Addiction Substance Abuse at Columbia University. (2011). *Adolescent substance use: America's #1 public health problem.* Retrieved from http://www.casacolumbia.org/addiction-research/reports/adolescent-substance-use

National Institute of Alcohol Use and Alcoholism. (2004). *Recommended alcohol questions*. Retrieved from http://www.niaaa.nih.gov/research/guidelines-and-resources/recommended-alcohol-questions

National Institute on Drug Abuse. (2014). Data harmonization projects. Retrieved from http://www.drugabuse.gov/researchers/research-resources/data-harmonization-projects

Teichman, M., Barnea, Z., & Ravav, G. (1989). Personality and substance use among adolescents: A longitudinal study. *British Journal of Addiction, 84*(2), 181-190. doi:10.1111/j.1360-0443.1989.tb00567.x

Villarosa, M. C., Madson, M. B., Zeigler-Hill, V., Noble, J. J., & Mohn, R. S. (2014). Social anxiety symptoms and drinking behaviors among college students: The mediating effects of drinking motives. *Psychology Of Addictive Behaviors*, *28*(3), 710-718. doi:10.1037/a0036501

Winters, K. (2003). Assessment of alcohol and other drug use behaviours among adolescents. In: Allen, J.P., Wilson, V.B. (Eds.), *Assessing Alcohol Problems* 2nd Ed, pp. 101-123. NIH Publication No. 03-3745. National Institute on Alcohol Abuse and Alcoholism, Bethesda.

Zucker RA. (2006). Alcohol use and the alcohol use disorders: A developmental-biopsychosocial systems formulation covering the life course. In: Cicchetti, D., Cohen, D.J. (Eds.), *Developmental Psychopathology, Risk, Disorder, and Adaptation (Volume 3)* (620-656). Hoboken, NJ: John Wiley & Sons Inc.

Table 1

Descriptive Statistics of all Primary Variables

	Mean	Standard Deviation	Range
Lifetime use Model	0.69	0.46	0-1
Proportion Model	0.22	0.20	0-1
Factor score Model	0	0.99	-1.17 – 3.01
GPA	2.82	0.76	1-4
Internalizing Symptomatology	0.56	0.39	0-3 (2.63)
Externalizing Symptomatology	0.26	0.32	0-3 (2.8)
Age	15.98	1.75	12-21
Gender	0.48	0.50	0-1 (0 = male)
Race	0.33	0.47	0-1 (1=minority)

Note: $N = 5{,}857$

Table 2

Correlations between all Primary Variables

Measure	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1. Lifetime use Model	___								
2. Proportion Model	.72**	---							
3. Factor score Model	.78**	.98**	---						
4. GPA	-.20**	-.23**	-.24**	---					
5. Internalizing Symptomatology	.18**	.23**	.24**	-.23**	___				
6. Externalizing Symptomatology	.31**	.47**	.47**	-.24**	.25**	___			
7. Age	-.24**	-.25**	-.26**	.09**	-.11**	.03	---		
8. Gender	.03*	.06**	.04*	.16**	-.12**	.14**	-.04*	---	
9. Race	-.08**	-.13**	-.12**	-.10**	.11**	.03	-.01	-.02	---

Note. *$p < .05$ **$p < .0001$

Table 3

Summary of Regression Analyses Predicting Academic Achievement (N = 5857)

Model		**Unstandardized B**	**SE(B)**	**Squared semi-partial correlation**	***t***	***df***	**F**	**Sig. (*p*)**	R^2
Control Variables	Overall model					3, 5853	84.02	<.0001	.04
	Intercept	-65.44	11.01		-5.94			< .0001	
	Gender	-.02	.02	.02	-12.23			< .0001	
	Age	.03	.01	.01	6.22			< .0001	
	Race	-.16	.02	0	-7.67			< .0001	
Lifetime Use	Overall model					4, 5852	122.54	<.0001	.08
	Intercept	-25.33	11.12		-2.28			< .05	
	Gender	-.23	.02	.02	-12.18			< .0001	
	Age	.01	.01	0	2.57			< .05	
	Race	-.19	.02	.01	-9.10			< .0001	
	Substance Use	**-.32**	**.02**	**.04**	**-15.11**			**< .0001**	
Proportion	Overall model					4, 5852	147.08	<.0001	.09
	Intercept	-16.00	11.07		-1.45			.15	
	Gender	-.22	.02	.02	-11.73			< .0001	
	Age	.01	.01	0	1.73			.08	
	Race	-.21	.02	.02	-10.28			< .0001	
	Substance Use	**-.87**	**.05**	**.05**	**-17.96**			**< .0001**	
Factor Score	Overall model					4, 5852	152.71	<.0001	.09
	Intercept	-12.21	11.08		-1.10			.27	
	Gender	-.23	.02	.02	-12.09			< .0001	
	Age	.01	.01	0	1.37			.17	
	Race	-.21	.02	.02	-10.18			< .0001	
	Substance Use	**-.18**	**.01**	**.05**	**-18.55**			**< .0001**	

Note: Substance use information is simply bolded for easy reference.

Table 4

Summary of Regression Analyses Predicting Internalizing Symptomatology (N = 5857)

Model		**Unstandardized B**	**SE(B)**	**Squared semi-partial correlation**	***t***	***df***	**F**	**Sig. (*p*)**	R^2
Control Variables	Overall model					3, 5853	79.98	<.0001	.04
	Intercept	52.77	5.65		9.34			< .0001	
	Gender	-.09	.01	.01	-9.38			< .0001	
	Age	-.03	0	.01	-9.24			< .0001	
	Race	.09	.01	0	8.32			< .0001	
Lifetime Use	Overall model					4, 5852	107.57	<.0001	.07
	Intercept	34.28	5.73		5.98			< .0001	
	Gender	-.10	.01	.02	-9.78			< .0001	
	Age	-.02	0	.01	-5.90			< .0001	
	Race	.10	.01	.01	9.59			< .0001	
	Substance Use	**.15**	**.01**	**.03**	**13.52**			**< .0001**	
Proportion	Overall model					4, 5852	149.49	<.0001	.09
	Intercept	26.61	5.67		4.69			< .0001	
	Gender	-.10	.01	.02	-10.50			< .0001	
	Age	-.01	0	0	-4.61			< .0001	
	Race	.12	.01	.02	11.04			< .0001	
	Substance Use	**.46**	**.03**	**.05**	**18.55**			**< .0001**	
Factor Score	Overall model					4, 5852	151.80	<.0001	.09
	Intercept	25.12	5.68		4.42			< .0001	
	Gender	-.10	.01	.02	-10.16			< .0001	
	Age	-.01	0	0	-4.32			< .0001	
	Race	.11	.01	.02	10.88			< .0001	
	Substance Use	**.10**	**.01**	**.05**	**18.78**			**< .0001**	

Note: Substance use information is simply bolded for easy reference.

Table 5

Summary of Regression Analyses Predicting Externalizing Symptomatology (N = 5857)

Model		**Unstandardized B**	**SE(B)**	**Squared semi-partial correlation**	***t***	***df***	**F**	**Sig. (*p*)**	R^2
Control Variables	Overall model					3, 5853	40.62	<.0001	.02
	Intercept	2.48	4.67		.53			.60	
	Gender	.09	.01	.02	10.82			< .0001	
	Age	0	0	0	-.48			.63	
	Race	.02	.01	0	2.12			< .05	
Lifetime Use	Overall model					4, 5852	202.04	<.0001	.12
	Intercept	-25.70	4.55		-5.64			< .0001	
	Gender	.09	.01	.02	10.93			< .0001	
	Age	.01	0	0	5.66			< .0001	
	Race	.04	0	0	4.49			< .0001	
	Substance Use	**.23**	**.01**	**.10**	**25.93**			**< .0001**	
Proportion	Overall model					4, 5852	499.96	<.0001	.25
	Intercept	-42.39	4.20		-10.08			< .0001	
	Gender	.08	.01	.01	10.42			< .0001	
	Age	.02	0	.01	10.09			< .0001	
	Race	.06	.01	.01	8.34			< .0001	
	Substance Use	**.79**	**.02**	**.23**	**42.89**			**< .0001**	
Factor Score	Overall model					4, 5852	495.41	<.0001	.25
	Intercept	-44.20	4.22		-10.47			< .0001	
	Gender	.08	.01	.02	11.25			< .0001	
	Age	.02	0	.01	10.52			< .0001	
	Race	.06	.01	.01	7.83			< .0001	
	Substance Use	**.16**	**0**	**.23**	**42.68**			**< .0001**	

Note: Substance use information is simply bolded for easy reference.

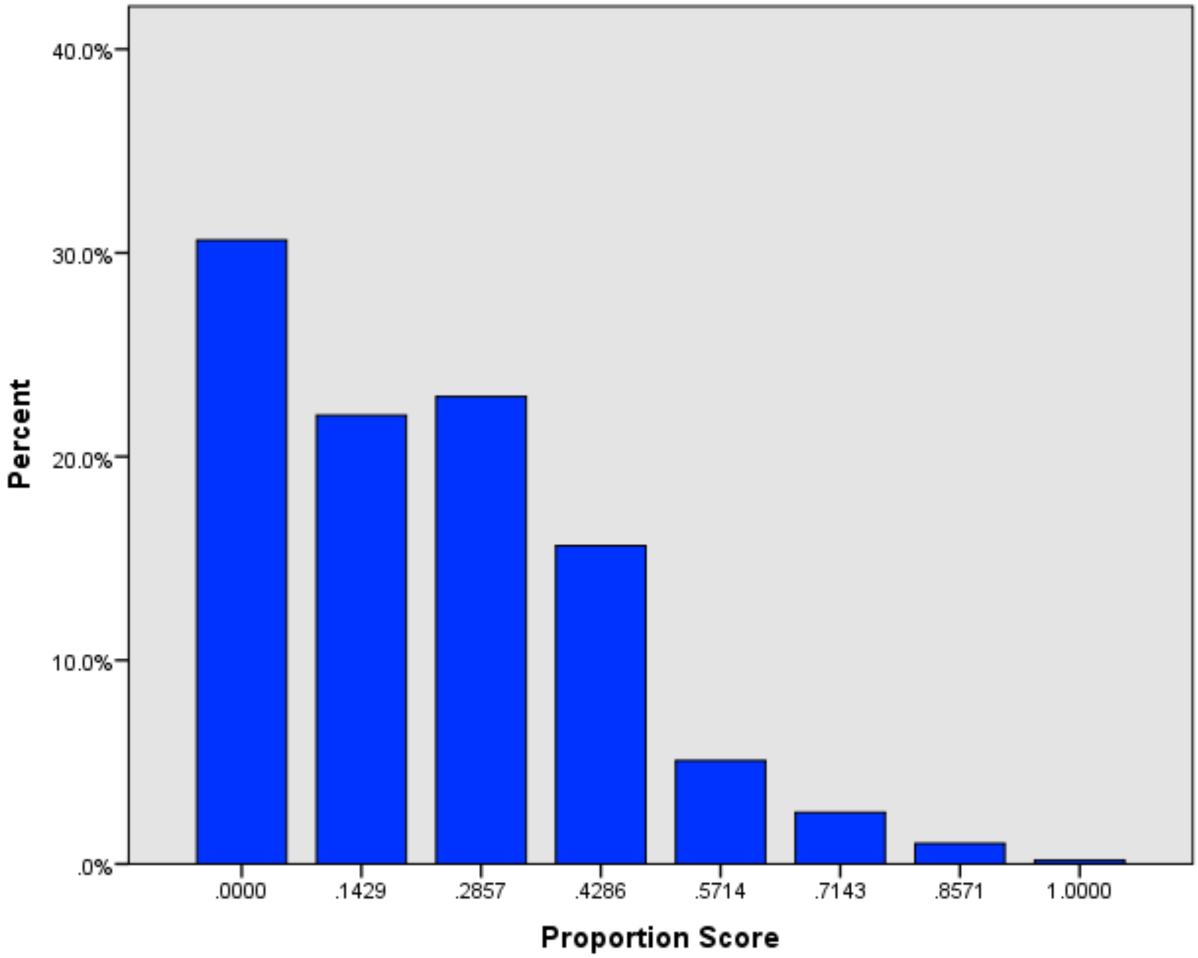
Figure Captions

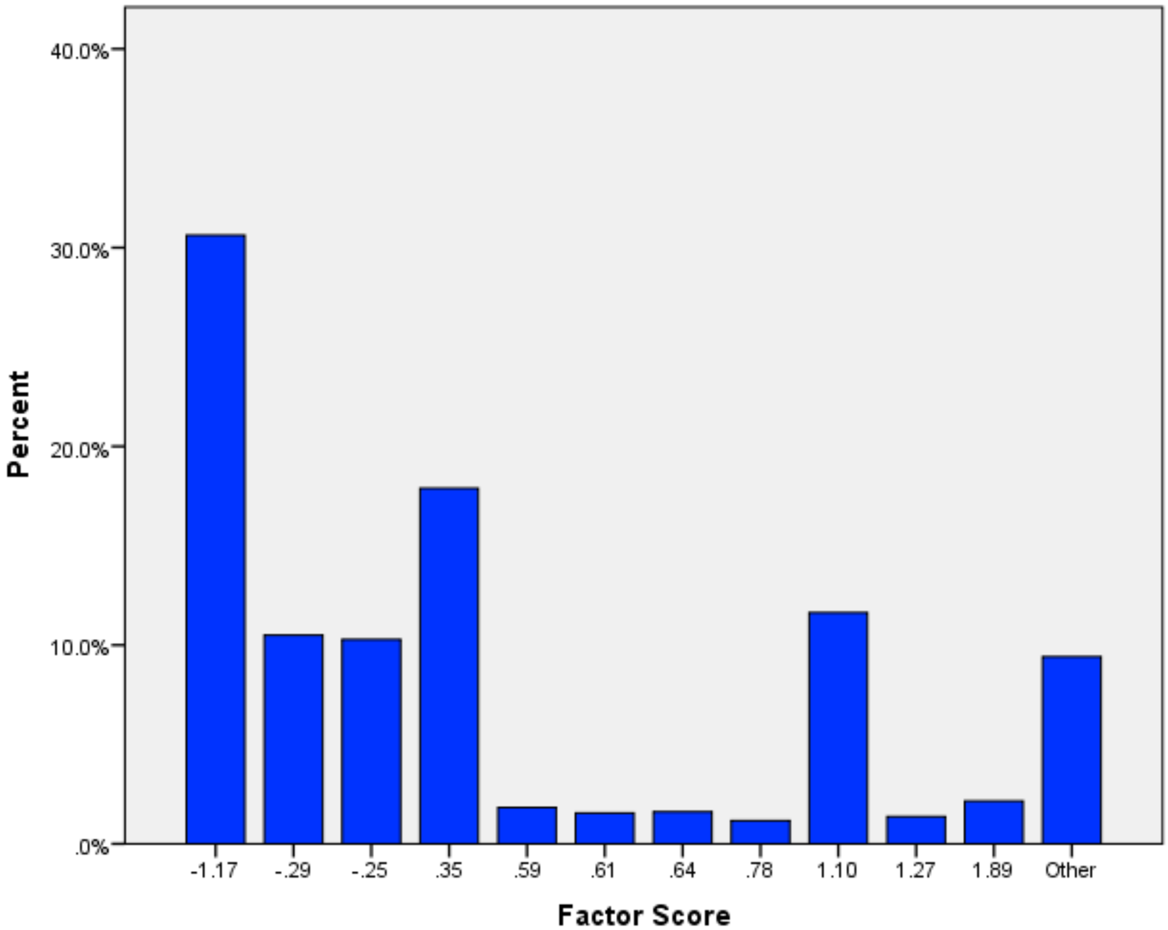
Figure 1. Distribution of Scores from the Proportion Model

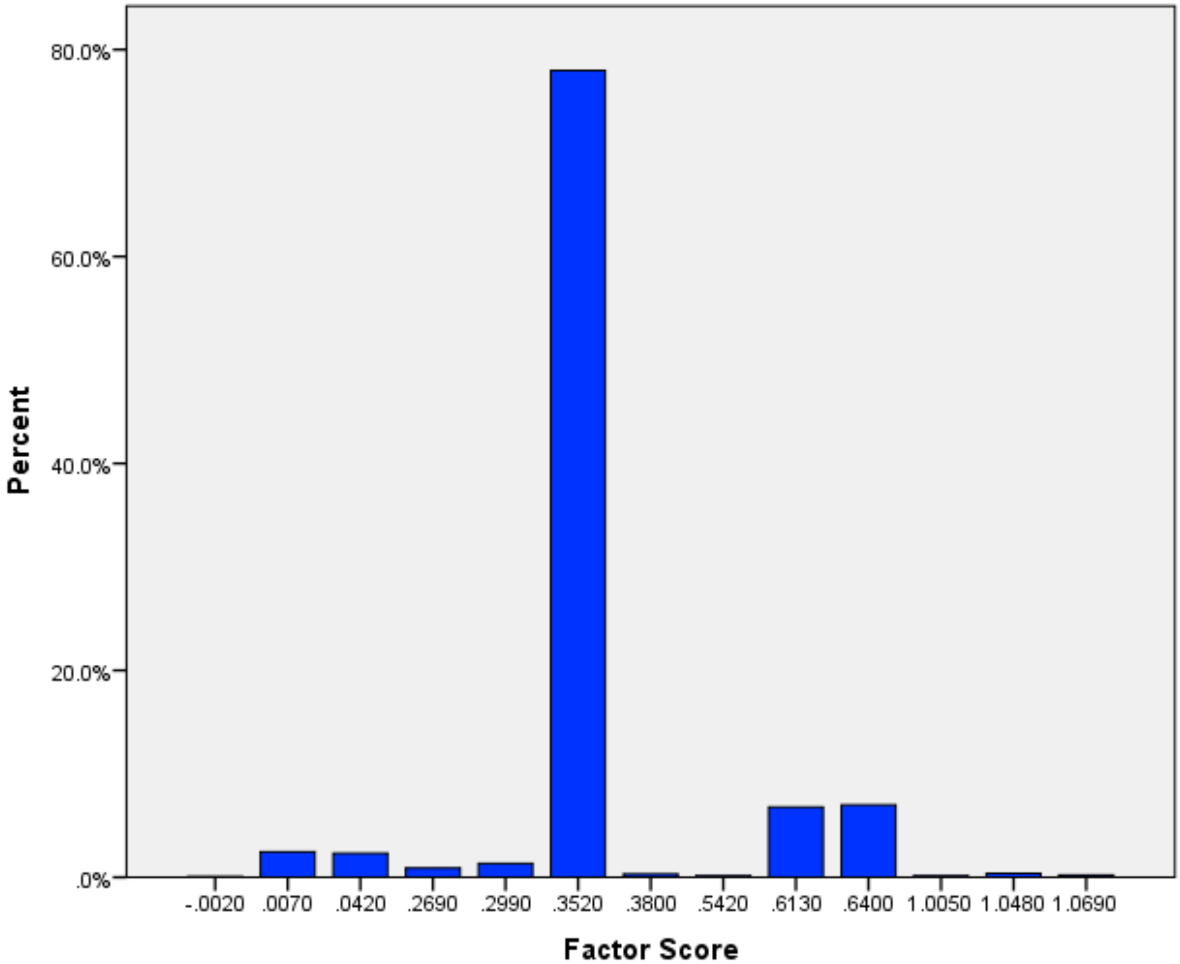
Figure 2. Distribution of Scores from the Factor Score Model

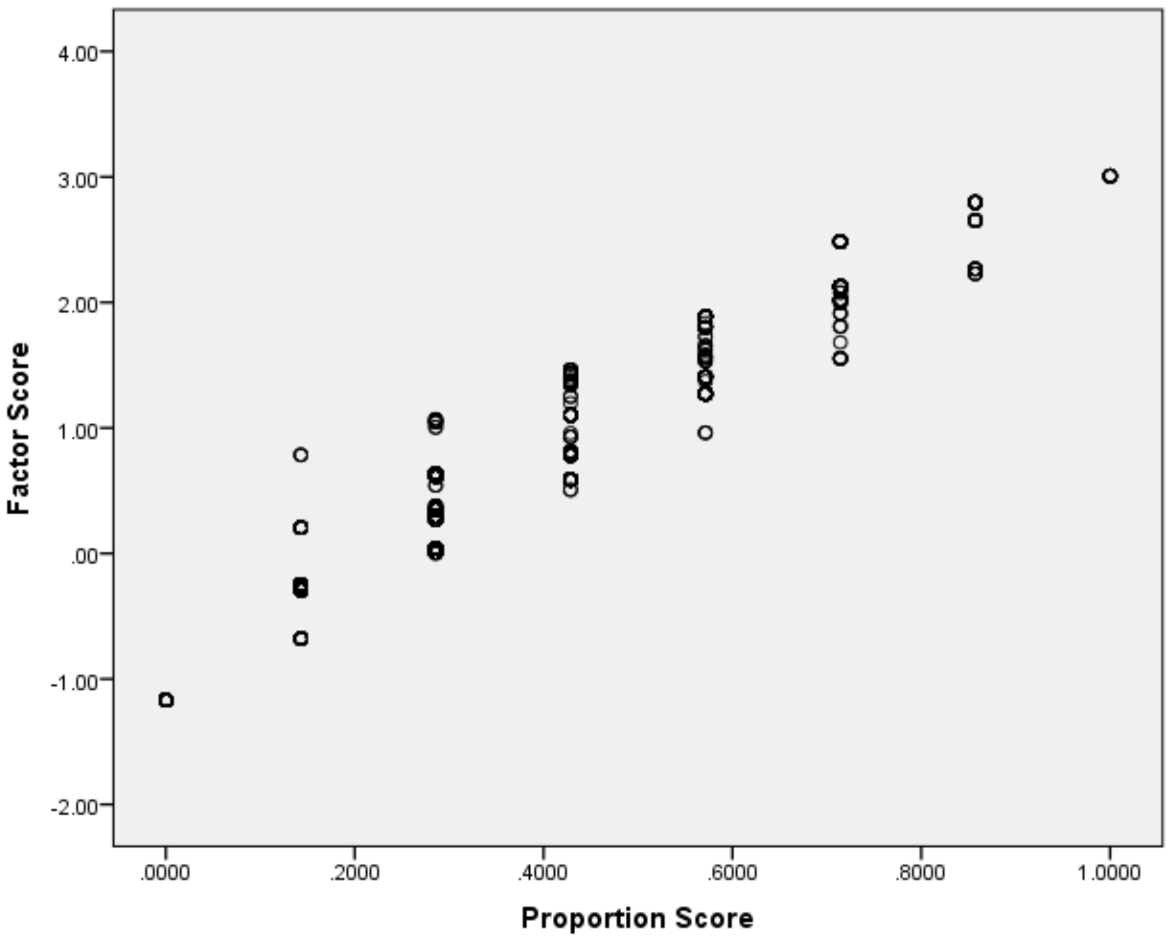
Figure 3. Distribution of Factor Scores for Participants with a Score of 2/7 on the Proportion Model

Figure 4. Bivariate Distribution between Proportion Model Scores and Factor Score Model Scores


40.0%
30.0%
20.0%
10.0%
.0%
Percent
.0000
.1429
.2857
.4286
.5714
.7143
.8571
1.0000
Proportion Score


40.0%
30.0%
20.0%
10.0%
.0%
Percent
-1.17
-.29
-.25
.35
.59
.61
.64
.78
1.10
1.27
1.89
Other
Factor Score


80.0%
60.0%
40.0%
20.0%
.0%
Percent
-.0020
.0070
.0420
.2690
.2990
.3520
.3800
.5420
.6130
.6400
1.0050
1.0480
1.0690
Factor Score


4.00
3.00
2.00
1.00
.00
-1.00
-2.00
Factor Score
.0000
.2000
.4000
.6000
.8000
1.0000
Proportion Score

Appendix

Selected Substance Use Items

Item	Response Options
Have you ever tried cigarette smoking, even just 1 or 2 puffs?	No Yes
How old were you when you used chewing tobacco or snuff for the first time?	Under one year 1-17 (by increments of 1) 18 years and older Never used chewing tobacco/snuff
Have you had a drink of beer, wine, or liquor-not just as sip or a taste of someone else's drink-more than 2 or 3 times in your life?	No Yes
During your life, how many times have you used marijuana?	1-150 (by varying increments) Never tried marijuana
During your life, how many times have you used cocaine?	1-50 (by varying increments) Never tried cocaine
During your life, how many times have you used inhalants, such as glue or solvents?	1-20 (by varying increments) Refused Never tried inhalants such as these
During your life, how many times have you used any of these types of illegal drugs (see question 14)?	1-150 (varying increments) Never tried any other type of illegal drug)

Academics and Education Items

Item	Response Options
During this school year…	
At the [most recent grading period/last grading period in the spring], what was your grade in English or language arts?	A B C D or lower
And what was your grade in mathematics?	
And what was your grade in history or social studies?	
And what was your grade in science?	

Note. Response options are the same for all items.

Feelings Scale

Item	Response Options
You were bothered by things that usually don't bother you.	Never/rarely (0) Sometimes (1) A lot of the time (2) Most/all the time (3)
You didn't feel like eating, your appetite was poor.	
You felt that you could not shake off the blues, even with help from you family and friends.	
You felt that you were just as good as other people. (R)	
You had trouble keeping your mind on what you were doing.	
You felt depressed.	
You felt that you were too tired to do things.	
You felt hopeful about the future. (R)	
You thought your life had been a failure.	
You felt fearful.	
You were happy. (R)	
You talked less than usual.	
You felt lonely.	
People were unfriendly to you.	
You enjoyed life. (R)	
You felt sad.	
You felt that people disliked you.	
It was hard to get started doing things.	

You felt like was not worth living.

Note. Response options are the same for all items.
Note. (R) indicates an item that was reverse-coded.

Delinquency Scale

Item	Response Options
In the past 12 months, how often did you paint graffiti or sign on someone else's property or in a public place?	Never 1 or 2 times 3 or 4 times 5 or more times
In the past 12 months, how often did you deliberately damage property that didn't belong to you?	
In the past 12 months, how often did you lie to your parents or guardians about where you had been or whom you were with?	
How often did you take something from a store without paying for it?	
How often did you get into a serious fight?	
How often did you hurt someone badly enough to need bandages or care from a doctor or nurse?	
How often did you run away from home?	
How often did you drive a car without its owner's permission?	
In the past 12 months, how often did you steal something worth more than $50?	
How often did you go into a house or building to steal something?	
How often did you use or threaten to use a weapon to get something from someone?	
How often did you sell marijuana or other drugs?	
How often did you steal something worth less than $50?	
In the past 12 months, how often did you take part in a fight where a group of your friends was against another group?	
How often were you loud, rowdy, or unruly in a public place?	

Note. Response options are the same for all items.