Drug and Alcohol Dependence 68 (2002) S7–S20

DRUG and ALCOHOL DEPENDENCE

www.elsevier.com/locate/drugalcdep

Assessing reliability of categorical substance use measures with latent class analysis

Brian P. Flaherty *

The Pennsylvania State University, S-159 Henderson Building, University Park, PA 16802, USA

Received 9 April 2002; received in revised form 17 June 2002; accepted 24 June 2002

Abstract

This article illustrates the use of the latent class model to identify classes of individuals and to assess the psychometric reliability of categorical items. The latent class model is a categorical latent variable model used to identify homogeneous classes of respondents such that class membership accounts for item responses. The assessment of measurement reliability comes directly from the estimates of the model. Although not based on classical test theory, the reliability assessment procedures described here answer the same question—that is, how consistent or dependable is measurement? The goal is to identify reliable indicators of a characteristic by examining measurement error and the inter-relatedness of the items. Methods for estimating the reliability of individual items as well as sets of items are presented. These methods are illustrated with data on cigarette smoking from a national sample of adolescents. By using the procedures described here, researchers are able to determine: (1) which classes of people are measured well and which are not; (2) which items perform well and which do not; and (3) whether items need to be altered or added in order to measure and identify particular classes better. © 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Adolescents; Categorical data; Cigarette smoking; Latent class analysis; Measurement reliability; Psychometrics

1. Introduction

Substance use data often include measurement error, sometimes in substantial amounts (Maisto et al., 1990; Cox et al., 1992; Harrison and Hughes, 1997; Golub et al., 2000; Biemer and Wiesen, 2002). Measurement error can affect categorical variables just as it can affect continuous variables. When items are categorical, measurement error results in the selection of an incorrect response category = for example, a cigarette smoker indicating no past smoking. As the amount of measurement error increases, data become less useful for theory development and model testing. Therefore, assessing random measurement error is an important step in any item and scale development. This is a primary goal of reliability analysis.

This article presents a method to examine the reliability of categorical items using the latent class model. The latent class model is a statistical model for categorical data that can be used to identify classes of respondents and estimate measurement error in categorical variables. Many theories and models in substance use research involve classes of people or stages of substance use. Examples include: stages of licit and illicit substance use (Kandel et al., 1987; Graham et al., 1991; Kandel et al., 1992; Kane and Yacoubian, 1999; Golub and Johnson, 2001); stages of adolescent alcohol use (Guo et al., 2000); types of alcoholics (Peters, 1997); and patterns of adult smoking (Shiffman et al., 1994). Results from a latent class analysis may be used for reliability assessment.

These procedures are useful for evaluating existing categorical items and for designing new ones. However, before reliability and the latent class model are presented, a model of the onset of nicotine dependence is introduced. This onset model is used to provide a context for subsequent discussion and it is also the basis of the empirical example.

* Tel.: +1-814-863-5836; fax: +1-814-863-0000
E-mail address: bxf4@psu.edu (B.P. Flaherty).

03765-8716/02/$ - see front matter
PII: S0376-8716(02)00210-7

1.1. Onset model of nicotine dependence

The development of nicotine addiction has been described as a sequence of stages distinguished by level and pattern of consumption (Leventhal and Cleary, 1980; Flay, 1993; Mayhew et al., 2000). The stages are contemplation, trying, experimentation, regular use, and nicotine dependence. The contemplation stage encompasses a period of time when a person is considering tobacco use, but has not yet tried it. The distinction between the trying stage and the experimentation stage is that trying includes the first few experiences with nicotine whereas experimentation is conceptualized as a period of irregular, but repeated use of tobacco. Regular use is a stage characterised by regular, possibly daily, tobacco use. This level of consumption may then lead to nicotine dependence.

This model of onset does not include other possible patterns of tobacco use (Darling and Cumsille, 2002). For example, an individual who has successfully quit does not fit into a stage in the model of onset. Someone who tried one cigarette 3 years ago and has not smoked since is no longer trying cigarette smoking. Thus, a model of interest may not apply to everyone in a sample and additional classes may be required to describe the sample.

This article uses the latent class model to identify classes of adolescents based on three items measuring cigarette smoking behavior and one item measuring beliefs about future smoking. Classes corresponding to stages of nicotine dependence onset are expected, but additional classes of use and future beliefs are also anticipated. With these goals in mind, reliability and the latent class model will be introduced. Measurement error in categorical variables is formalized in terms of the latent class model, with the theory of nicotine dependence onset used as an example throughout.

1.2. Reliability

Assessing the psychometric properties of measurement instruments is an important part of theory development. Typically, this assessment is divided into two separate but related tasks: reliability and validity assessment. Reliability concerns the ability to obtain repeatable or consistent measurements (Cronbach, 1951; Guilford, 1954; Lord and Novick, 1968; Carmines and Zeller, 1979). Validity is the degree to which behaviors of interest are truly being measured (Cook and Campbell, 1979). A measure can only be as valid as it is reliable (Kerlinger, 1964; Lord and Novick, 1968). Therefore, reliable measurement is a critical aspect of theory development.

Much work on reliability in psychometric theory has been based on classical test theory (CTT; Lord and Novick, 1968). In the CTT framework, an individual's observed score is the sum of two parts: that person's true score and an error score. The true score is conceived of as the score that would be observed if there is no random measurement error. That is, if a measurement process was unaffected by random error, the measurements would be perfectly reliable. To the extent that random measurement error affects the measurement process, reliability decreases.

Along with the CTT model of test scores, the psychometric theory of reliability also includes the use of multiple indicators (Spearman, 1910; Kuder and Richardson, 1937; Cronbach, 1951). The term 'multiple indicators' refers to the use of multiple items to measure a behavior or characteristic that cannot be measured directly. These unobservable characteristics are often referred to as constructs. Any single indicator is an imperfect measure of such a construct. However, several indicators that each measure a construct of interest provide a more precise assessment of the construct. One result of using multiple indicators is that one can estimate random measurement error and, therefore, also estimate reliability. If all observed scores are sums of true scores and uncorrelated random errors, then multiple items that each measure the same construct will covary to the extent that the true scores are expressed. In other words, if items are individually reliable and all reflect the same construct of interest, these items will be highly correlated. This implies high reliability of the set of items or scale. As correlations among the items decrease due to random measurement error, scale reliability likewise decreases. When only one indicator of a construct is used, measurement error and reliability cannot be estimated because there are no relations among items to examine. Therefore, the assessment of reliability requires multiple indicators. Probably the best known CTT index of reliability for continuous measures is Cronbach's α (Cronbach, 1951).

The concepts and goals of reliability assessment are the same in CTT and the work presented here. This work is not based on CTT, but it is based on the concepts underlying CTT that have developed over the past several decades. The purpose of the procedures presented here is to develop reliable measurement, by assessing random measurement error among multiple indicators of a latent construct.

1.3. Why assess reliability with latent class analysis

Before introducing the latent class model, it is fair to ask why assess reliability with latent class analysis when there is existing psychometric theory? This approach is beneficial for at least two reasons. First, existing methods like Cronbach's α and the common factor model (Gorsuch, 1983) assume that the construct of interest is represented by a single continuous dimension and that individuals may be ordered along that dimension. However, this may not always be desirable or appropriate. Consider a latent class model of cigarette smoking onset with three classes; (1) Never smoked; (2) Trying/experimenting; and (3) Regular use. An underlying continuum in this case is not problematic; one could imagine a latent continuum of lifetime use accounting for observations. But, if there is also a latent class of quitters or past smokers, the continuum breaks down. A latent class of past smokers does not fit along the continuum. Past smokers could be distributed over a lifetime use continuum because some may have smoked relatively few cigarettes and others may have smoked many prior to quitting. The key feature of the past smoker class, and one that a single continuum ignores, is the pattern of responses indicating past but not recent use.

Building further on this notion of a dimension, reliability coefficients like Cronbach's α and the common factor model are assumed to hold across the sample and across all levels of the construct being measured (Carmines and Zeller, 1979). This assumption implies that the sample is drawn from a single population. However, an implicit assumption of the latent class model is that there are subpopulations. From the outset, it is assumed that the sample is comprised of people from different classes and that, across the classes, item–latent variable relations may differ. If there are subgroups in a data set, then methods based on an underlying continuous dimension may be inappropriate.

The second reason latent class analysis should be preferred to CTT and factor analysis in certain circumstances concerns distributional assumptions. The few assumptions made in the latent class model are also made in the common factor model. These shared assumptions are the independence of sampled individuals and the conditional independence of items given the latent variable (Bartholomew, 1987; also see Section 1.4.1). However, the factor model also assumes that the latent variable and residual errors are normally distributed. These assumptions are clearly inappropriate for categorical variables. For the cigarette smoking example, it is impossible for lifetime cigarette consumption to be normally distributed. Also, residual errors among categorical items cannot be normally distributed. In contrast, the latent class model has no distributional assumptions on either the latent class variable or the errors.

1.4. Latent class analysis

Consider four items concerning cigarette smoking. Two of the items measure lifetime cigarette smoking and are considered indicators of regular cigarette smoking: 'Have you ever smoked 100 or more cigarettes?' (1 = 'No'; 2 = 'Yes') and 'Have you ever smoked every day for a month?' (1 = 'No'; 2 = 'Yes'). The third item concerns recent use, 'On how many of the past 30 days have you smoked?' ('1' = None; '2' = one or more, but less than 30; '3' = all 30). The fourth item concerns beliefs about future smoking, 'Do you think you will be smoking 1 year from now? ('1' = Definitely not; '2' = Probably not; '3' = Probably or definitely will).

If a sample of adolescents with past smoking experience are asked these four questions, one would expect relations among the responses. For example, if a person indicates that he or she has smoked every day for a month, it is more likely that this individual has also smoked in the past 30 days. It is reasonable to think that someone's responses to items about cigarette smoking are strongly related to that person's smoking behavior. These relations among items are the basis of latent class analysis. The latent class model identifies classes of people such that class membership accounts for observed responses.

In latent class analysis, the unit of analysis is the response pattern. A response pattern is the set of responses given by an individual to a set of items. In the example, there are four items with a total of 36 possible response patterns ($2 \times 2 \times 3 \times 3$). Table 1 contains some possible response patterns us for the four cigarette smoking items.

Suppose previous work indicates that two groups of smokers are expected: experimenters and regular users. Given this expected class structure, one can go through the response patterns in Table 1 and indicate the expected latent class membership for people making each response pattern. (Recall that this sample consists of adolescents who report some cigarette smoking experience.) Someone who responds that he or she has not smoked over 100 cigarettes, has never smoked every day for a month, has not smoked at all in the past 30 days and believes that he or she definitely will not smoke in 1 year is most likely a member of the tried/experimenter latent class. This response pattern is the first in Table 1. On the other hand, someone who responds 'Yes' to the first two items, indicates smoking

Table 1
A subset of the possible response patterns for the four cigarette smoking items asked of a sample of self-identified adolescent smokers

Response pattern	Expected probability in tried or experimenter group	Expected probability in regular use group
1111	High	Low
1122	High	Low
1132[a]	??	??
1133[a]	??	??
2112	??	??
2123	< 0.50	> 0.50
1233	< 0.50	> 0.50
2123	Low	High
2233	Low	High

[a] An illogical or inconsistent response pattern.

every day in the past month, and indicates that he or she will be smoking in 1 year is most likely in the 'regular smoker' latent class. This is the last response pattern shown in Table 1.

Table 1 also contains two response patterns that are internally inconsistent. For example, someone who reports that he or she has not ever smoked every day for a month, but also reports having smoked on every day of the past 30 is inconsistent. It is worth considering for a moment how this would be handled in a manifest variable analysis. If respondents were being classified, how is an inconsistent response pattern handled? In this situation the researcher must either choose the class membership of this person, or drop the person from the analysis. Neither of these choices is satisfactory; uncertainty and obvious measurement error are ignored. An advantage of the latent class model is that both consistent and inconsistent response patterns are modeled and information in all the data is retained.

There is also an example of a response pattern in Table 1 that does not fit either class, but also is not internally inconsistent. Patterns like this may represent an additional latent class. The response pattern (2, 1, 1, 2) indicates that a person has smoked 100 or more cigarettes, has never smoked every day for a month, has not smoked in the past 30 days, and believes that he or she will probably not smoke in 1 year. This response pattern represents someone who did smoke regularly, does not now, and is uncertain about future smoking. There is nothing inherently inconsistent in this response pattern. If there are only a few response patterns like this in a sample, the pattern may be treated as error. However, if there are many people associated with this pattern, an additional latent class of people who no longer smoke might be required to model the data well.

1.4.1. The latent class model

Let $\mathbf{W} = (W_1, W_2, \ldots, W_p)$ denote a set of p manifest categorical variables (e.g. the cigarette smoking items) and let $\mathbf{d} = (d_1, d_2, \ldots, d_p)$ denote a vector of the number of response categories for each variable in $\mathbf{W}$ (bold notation denotes vectors). For the cigarette smoking items described in Section 1.4, $\mathbf{d} = (2, 2, 3, 3)$. Let $\mathbf{w} = (w_1, w_2, \ldots, w_p)$ denote a particular response pattern (e.g. a row in Table 1). Let L denote the latent class variable with C classes and let $P(\mathbf{W} = \mathbf{w}) = \pi_{\mathbf{w}}$ denote the probability of observing the response pattern $\mathbf{w}$. Then, the latent class model may be written:

$$P(\mathbf{W} = \mathbf{w}) = \pi_{\mathbf{w}} = \sum_{c=1}^{C} \gamma_c \prod_{i=1}^{p} \rho^{(i)}_{w_i|c} \tag{1}$$

where $\gamma_c = P(L = c)$ for example, the probability of being a member of the current smoker latent class and $\rho^{(i)}_{w_i|c}$ denotes the probability of response w_i to item W_i conditional on membership in latent class c, for example, the probability someone in the current smoker latent class says 'Yes' to the question about smoking 100 or more cigarettes.

Eq. (1) has two types of parameters: latent class probabilities (denoted by γ) and conditional response probabilities (denoted by ρ; a ρ without any super- or subscripts will denote a single, generic ρ parameter). Both the γ and ρ parameters are probabilities, and as such they range from 0 to 1. Parameters are quantities that describe a population. The parameters of the latent class model are commonly estimated with the expectation-maximization algorithm (EM; Goodman, 1974; Dempster et al., 1977). Parameter estimates are typically denoted with hat (^)—for example, $\hat{\gamma}_1$ is the estimated probability for Latent Class 1. In Eq. (1) γ_c represents the probability a randomly drawn person belongs to latent class c. The γ parameters are also the latent class proportions, that is the proportion of the population in each latent class. If regular smokers are the second latent class, $\hat{\gamma}_2$ is the estimated proportion of the sample in the regular smoker class. It is also the probability that a randomly drawn person is a member of the regular smoker latent class.

The conditional response probabilities are the within latent class distributions of the response categories. For example $\rho^{(4)}_{1|2}$ denotes the probability of a 'Definitely not' response (a '1') to item four conditional on membership in latent class 2 (e.g. regular smokers). These parameters are the basis of interpretation in the latent class model. For example, a latent class with high probabilities of responding 'Yes' to the 100-cigarettes item and the smoke every day for a month item, a high probability of indicating smoking every day of the past 30, and a high probability of planning to smoke in the future can be interpreted as a regular smoker class.

A $\hat{\rho}$ equal or close to 1.0 signifies a behavior or characteristic that most or all members of the latent class endorse. (In the clinical literature, these items referred to as pathognomonic items—that is, they are very good diagnostic indicators.) On the other hand, $\hat{\rho}$'s near or equal to zero indicate that members of a latent class are not expected to endorse the respective response category. As the value of a $\hat{\rho}$ for Item W_j, approaches $1/d_j$, the item becomes less informative about that class. Essentially, a $\hat{\rho}$ near $1/d_j$ is random, that is, latent class membership is not predictive of item response.

Eq. (1) can be written as the simple product of the ρ's due to the assumption of conditional or local independence. A response to Item W_i is assumed to be independent of a response to Item W_j, $i \neq j$ conditional on latent class membership. In other words, a response to one item tells us nothing about a response to another item once we take latent class membership into account. One implication of this assumption is that the probability of a particular response pattern for a single latent class is the product of the individual response probabilities for that latent class. Another assumption of the latent class model is that the classes are mutually exclusive and exhaustive. Everyone is assumed to be a member of one and only one latent class.

1.4.2. Model fit and selection

The fit of a latent class model can be assessed with Pearson's X^2 or the likelihood ratio statistic G^2, just as in contingency table analysis. An X^2 or G^2 near or less than the degrees of freedom (df) is a rough indicator of a good model fit. The df for the latent class model are:

$$\mathrm{df} = \left(\prod_{j=1}^{p} d_j\right) - \left(C + \sum_{j=1}^{p} C(d_j - 1)\right) \quad (2)$$

The term on the left of Eq. (2) is the number of possible response patterns and the term on the right is the number of estimates in an unrestricted model plus one.

When the number of classes is unknown, choosing the number of classes becomes a model selection issue. That is, one must select the best model from a set of models with different numbers of classes. For technical reasons, latent class models with different numbers of classes cannot be compared via nested model teeing (Aitkin and Rubin, 1985). One frequently used approach is to fit models with different numbers of latent classes and choose the one that looks the best, that is, the model that fits the data and is interpretable.

Another approach to choosing among models with different numbers of classes is to use an information criterion statistic, such as Akaike's Information Criterion (AIC; Sakamoto, 1991) or the Bayes Information Criterion (BIC; Raftery, 1995). These statistics balance parsimony and model complexity. Lower values of AIC and BIC indicate better fit, therefore, a model with the lowest AIC and/or BIC is generally preferred.

1.4.3. Parameter restrictions

Imposing parameter restrictions reduces the number of estimates in a model, thus making the model simpler. There are two common types of parameter restrictions in latent class models: fixing parameters to a particular value, and equality constraints. Fixing parameters means that one specifies a value for a particular parameter rather than estimating it. When a parameter is fixed, it does not use a degree of freedom for estimation. Equality constraints are used to cause a set of parameters to be estimated at the same value. This set counts as a single parameter with respect to the model degrees of freedom. Both types of restrictions are used in the analyses described below.

1.5. Reliability via latent class models

Clogg and Manning (1996) reported the first application of the latent class model to reliability analysis. In general, reliability with latent class models is operationalized as the predictive power a latent class has for an item response and vice versa. This is analogous to CTT in which the true score predicts the scale or item score. If an item response is a reliable indicator of latent class members of the class should make that response quite often. If the response in question is not consistently produced by members of the particular latent class, then it is an unreliable indicator. Similarly, a single item response can be highly predictive of a particular latent class. Item responses that do not strongly point to latent class membership are also not reliable.

Two types of reliability are presented by Clogg and Manning (1996): item-specific reliability and item-set reliability. Item-specific reliability assesses how reliably a single item measures a particular latent class. The most basic of the item-specific reliability indices are the $\hat{\rho}$'s. If for a particular latent class one response category of an item has a very high $\hat{\rho}$ and the other response categories have correspondingly low $\hat{\rho}$'s then that item is a reliable indicator of that latent class. An unreliable item is one where all the $\hat{\rho}$'s are essentially random (i.e. equal to $1/d_j$). A second measure of item-specific reliability is the predictive value an item response has for a particular latent class. Item response may be used to predict latent class as follows:

$$P(L = c | W_i = w_i) = \hat{\pi}_{L|W_i = w_i} = \frac{\hat{\gamma}_c \hat{\rho}^{(i)}_{w_i|L=c}}{\hat{\pi}_{W_i = w_i}}, \quad (3)$$

where $\hat{\pi}_{W_i = w_i}$ denotes the estimated marginal probability of response w_i to Item W_i. These quantities are the reverse of the ρ parameters. Rather than starting with latent class membership and predicting item response, estimates based on Eq. (3) convey how well a single item response discriminates among the latent classes.

Odds ratios can also be used to assess reliability in the latent class model:

$$\hat{\psi} = \frac{(\hat{\rho}^{(i)}_{w_i=1|L=1})(\hat{\rho}^{(i)}_{w_i=2|L=2})}{(\hat{\rho}^{(i)}_{w_i=2|L=1})(\hat{\rho}^{(i)}_{w_i=1|L=2})}. \quad (4)$$

Eq. (4) compares the probability of a 1 or 2 response to Item W_i given membership in either Latent Class 1 or 2. It conveys how well a pair of latent classes distinguished a pair of item responses. If $\hat{\psi} = 1$, then there is no relation between this pair of latent classes and this pair of item responses. The odds ratio ($\hat{\psi}$) can also be transformed to a correlation metric with Yule's Q transform (Agresti, 1990):

$$\hat{Q} = \frac{\hat{\psi} - 1}{\hat{\psi} + 1}.$$

These two quantities offer the advantage of summarizing the relations across two classes and two response categories simultaneously. Note that when a ρ parameter is estimated at or fixed to 0.0, $\hat{\psi}$ is not useful because it is always zero or infinity.

The final type of reliability index discussed by Clogg and Manning is item = set reliability. Rather than studying the items individually, one can examine how well a group of items measures the latent classes, just as one would examine how well a scale measures something like self-esteem. As discussed in Section 1.2, individual items are limited in the capability to measure latent variables. Also, items may distinguish certain classes, but not others. The items as a set should distinguish that the latent classes well. An individual's response pattern ($\mathbf{W}=\mathbf{w}$) is the analogue to a continuous variable scale score. Item-set reliability is:

$$\hat{\pi}_{L=c|\mathbf{W}=\mathbf{w}} = \frac{\hat{\pi}_{L=c,\mathbf{W}=\mathbf{w}}}{\hat{\pi}_{\mathbf{W}=\mathbf{w}}}$$

where $\pi_{L=c,\mathbf{W}=\mathbf{w}}$ is the unconditional probability of both membership in latent class c and response pattern $\mathbf{w}$, and $\pi_{\mathbf{W}=\mathbf{w}}$ the marginal probability of response pattern $\mathbf{w}$. These quantities are the probability of latent class membership conditional on an entire response pattern—that is, how well do the response patterns distinguish the latent classes?

These five indices provide different views of reliability, but all are directed at assessing how strongly the latent classes and items are related. All these quantities are easily obtained directly from the results of a latent class.

2. Methods

2.1. Data

Data are from the 1989 Teenage Attitudes and Practices Survey (TAPS; Allen et al., 1992; Centers for Disease Control and Prevention, 1991; Moss et al., 1992). The TAPS survey was developed by the Office of Smoking and Health at the US Centers for Disease Control in order to monitor adolescent smoking behavior and attitudes nationally (Allen et al., 1992).

Adolescent tobacco use and attitudes are important aspects of the model of nicotine dependence onset, discussed in Section 1.1. Therefore, the TAPS data would seem ideal for examining the measurement of the stages in the onset model of nicotine dependence. In fact, prevalences of classes of tobacco users were one of the first results reported from these data (Moss et al., 1992). However, the TAPS survey was not designed with multiple indicator latent variable models in mind. A problem with using the TAPS data for this work is that, every respondent did not answer every relevant question. If someone indicated that he or she did not ever try a cigarette that person was skipped out of the bulk of questions about cigarette smoking. Although this minimizes respondent burden, it also precludes the ability to estimate measurement error and reliability across the sample. When only individuals who said ‘Yes’ to a particular item are included, the reliability of that item cannot be estimated—there is no one who said ‘No’ to the item in the sub-sample. Therefore, in TAPS, one can only examine measurement reliability within subgroups of respondents and for particular items.

In 1989, 9965 individuals were surveyed. Telephone interviews were conducted with 9135 people (92%). Those who could not be reached by telephone, 830 (8%), were sent a mail survey (Moss et al., 1992). In order to preclude potential differences due to survey administration mode (Turner et al., 1992), mail survey recipients are not included in these analyses. The sample was further restricted to include only those individuals who reported having tried cigarettes, 3106 adolescents. Of these 3106 respondents, 53% are male. Eighty-nine percent identified themselves as ‘white,’ 8% as ‘black,’ and 3% endorsed the ‘Other’ category. The age range is 12–18 years old, with a mean of 15.9 years and standard deviation of 1.74 years.

2.2. Items

The four items listed in Section 1.4 are analyzed here. Rather than reiterate the list, only coding changes from the original survey are described. The third item, ‘On how many of the last 30 days did you smoke?’ was originally a count variable ranging from 0 to 30. However, the distribution of this variable was bimodal. There were 1720 ‘0’ responses indicating no smoking in the past 30 days, and 525 ‘30’ responses indicating that the respondent smoked every day of the past 30. The range between 0 and 30 had fewer people at each value. For this reason the item was recorded to form a three level variable: (1) No days in the past 30, (2) between 1 and 29 days, and (3) all 30 days in the past 30. The fourth item, ‘Do you think you will be smoking cigarettes 1 year from now?’ originally had four response categories: ‘Definitely not’, ‘Probably not’, ‘Probably yes’, and ‘Definitely yes’. The ‘Definitely yes’ category was rarely endorsed, so the two ‘Yes’ response categories were collapsed.

2.3. Analysis

Analyses were run with a UNIX version of the LTA program (Collins et al., 1999) maintained by the author. The reliability indices were calculated with a FORTRAN 90 program written by the author.

Table 2
Information relevant to model fit for unrestricted latent class models with two to five latent classes

Number of classes	Number of parameters	Degrees of freedom	G^2	P-value	AIC
2	13	22	765.57	< 0.00001	791.57
3	20	15	209.14	< 0.00001	249.14
4	27	8	21.38	< 0.0062	75.38
5	34	1	1.37	0.2411	69.37

The modeling approach taken here begins with an unconstrained latent class model in order to determine the appropriate number of classes (Clogg, 1995). Based on the model of nicotine addiction and on the limitations imposed by the restricted set of items, one would expect at least two classes in these data: experimenters and regular smokers. However, as mentioned earlier, an onset model may not account for everyone in a sample. Therefore, models with successively more classes were run until an interpretable model which fit the data was achieved.

After the number of classes was chosen, a series of more restrictive models were tested via nested model comparisons. Parameter restrictions were applied where it was substantively meaningful to fix a parameter to a particular value or to constrain a set of parameters to be equal. For example, any classes with low cigarette-smoking rates would be expected to have equally high probabilities of responding negatively to indicators of regular use. Any parameter restrictions that resulted in poor model fit were not retained.

These data include some missing values due to subject non-response. In addition, the first three items had “Don’t know” response categories that were treated as missing. The LTA program handles cases with missing data, so no cases were excluded due to missing data. However, use of this missing data procedure requires an assumption of ‘missing at random’ (Schafer, 1997). This assumption means that one is assuming that observed data can account for the missing data. With respect to the analyses here, it is assumed that the same latent class structure holds for people with complete data and incomplete data.

Also, the TAPS data are based on a complex sample design (Massey et al., 1989), but the sampling weights are not used here because the software does not incorporate sample weights. Any results of this study are restricted to the sample at hand, and do not necessarily generalize to the 1989 US population of adolescents.

3. Results

Table 2 contains information on a series of unrestricted latent class models including two to five classes. The fit of each model was assessed using the G^2, AIC, and BIC.

The null hypothesis being tested is the model being fit. Thus, when a model fits well the null hypothesis is not rejected. As seen in Table 2, the addition of each class results in the loss of 7 df. The five-class model has the only non-significant fit statistic. Furthermore, the five-class model also has the lowest AIC value. BIC values were similar and are not shown.

3.1. Model with parameter restrictions

Restrictions were applied to the model one or two at a time and the fit of the newly restricted model was compared with the fit of the model without the particular parameter restriction. Fourteen successively more restrictive models were fit, two of which were rejected due to poor fit. The final model had a $G^2 = 9.97$ with 16 df and AIC $= 47.97$. Table 3 contains the final $\hat{\gamma}'$s. Table 4 contains the final $\hat{\rho}'$s along with the final pattern of parameter restrictions.

The first class is made up of people who respond ‘No’ to the first two items; are most likely to indicate no days smoked in the past 30; and are among the most definitive in the belief that they will not be smoking in the next year. The first two item’s $\hat{\rho}$’s for the first latent class are examples of fixed parameter restrictions. The probability of a ‘No’ response to each of these items was fixed to 1.0[1]. This restriction means that no one in this class responded ‘Yes’ to either of the first two items. This class is labeled ‘past experimenters’.

Latent Class 2 are regular smokers. These individuals have the highest rates of endorsing ‘Yes’ to both past

Table 3
Estimated latent class membership probabilities ($\hat{\gamma}$) for the five-class model with restrictions

Latent class	1	2	3	4	5
$\hat{\gamma}$	0.56	0.19	0.08	0.06	0.10

[1] Algorithms like EM can encounter difficulties when estimates approach boundary values (here 0.0 and 1.0; see Schafer, 1997). In order to avoid these problems, several ρ parameters were fixed to 0.0 and 1.0 and model fit was not adversely effected.

Table 4
Conditional response probabilities ($\hat{\rho}$'s) for the restricted five-class model

Latent class	Ever smoke 100+cigarettes	Ever smoke every day for a month	Days smoked in past 30	Plan to smoke in 1 year
Response category 1				
	No	No	None	Def. Not
(1) Past experimenters	(1.000)	(1.000)	0.878	0.756 [9]
(2) Regular smokers	0.019 [1]	(0.000)	0.005	0.060
(3) Past smokers	0.353 [2]	0.353 [2]	0.745	0.756 [9]
(4) Irregular smokers	0.019 [1]	0.543	0.019 [7]	0.088 [10]
(5) Experimenters	0.906 [3]	0.906 [3]	0.019 [7]	0.088 [10]
Response category 2				
	Yes	Yes	Some	Poss. Not
(1) Past experimenters	(0.000)	(0.000)	0.122	0.227 [11]
(2) Regular smokers	0.981 [4]	(1.000)	0.103	0.214
(3) Past smokers	0.647 [5]	0.647 [5]	0.255	0.227 [11]
(4) Irregular smokers	0.981 [4]	0.457	0.981 [8]	0.558 [12]
(5) Experimenters	0.094 [6]	0.094 [6]	0.981 [8]	0.558 [12]
Response category 3				
			30	Yes
(1) Past experimenters	NA	NA	(0.000)	0.017 [13]
(2) Regular smokers	NA	NA	0.891	0.727
(3) Past smokers	NA	NA	(0.000)	0.017[13]
(4) Irregular smokers	NA	NA	(0.000)	0.354 [14]
(5) Experimenters	NA	NA	(0.000)	0.354 [14]

Fixed values are enclosed in parentheses and parameters constrained to be equal are denoted with identical superscripts. Values are provided to three decimal places because subsequent calculations require it.

smoking items. Members of this class are nearly certain to indicate having smoked 100+ cigarettes and having ever smoked every day for a month. Members of this class are most likely to indicate that they have smoked every day of the past 30. Finally, this class has the highest probability of indicating an affirmative belief regarding their smoking in 1 year.

The third latent class have or are trying to quit. Members of this class are about twice as likely to say 'Yes' as they are to say 'No' to the first two items. However, 75% of this group say they have not smoked at all in the past 30 days and no one in this class indicates smoking every day in the past 30. Also members of this group are as likely as past experimenters to respond 'Definitely not' to the future smoking item. This class can be thought of as 'past smokers'. Members indicate higher rates of past smoking with low rates of current smoking and strong plans not to smoke in 1 year.

Members of the fourth latent class are as likely as the regular smokers to indicate that they have smoked 100+ cigarettes. However, these individuals are nearly equally likely to indicate ever smoking every day for a month as not. A central feature of this class is the high probability of members reporting smoking on some days in the past 30, but not all 30 days. Also, after the regular smokers, this class has the highest probability of affirmative beliefs about smoking in the next year. Members of this class are labeled irregular smokers.

The fifth latent class are people in the experimentation phase. They have equally low probabilities of reporting ever smoking 100+ cigarettes or ever smoking every day for a month. These individuals are highly likely to report some smoking in the past 30 days. Furthermore, these individuals are also unsure about smoking in the next year.

3.2. Reliability of the restricted model

3.2.1. Single item reliability indices

Once a final latent class model that fits the data has been determined, reliability can be examined. The $\hat{\rho}$'s in Table 4 are variable across items and classes. For example, Items 1 and 2 appear to be very reliable indicators of Latent Classes 1, 2, and 5. Indeed, because the ρ parameters for items 1 and 2 in Latent Class 1 were able to be fixed to 1.0 and 0.0 without adversely effecting model fit, these two items are perfect indicators of Latent Class 1. But, Items 1 and 2 are both unreliable indicators of Latent Class 3. For Latent Class 4, Item 1 does well, but Item 2 does not.

Rather than present all the odds ratios ($\hat{\psi}$) and Yule's Q values, a few selected values are described. It is easy to see the comparison being made if a 2×2 table of the relevant $\hat{\rho}$'s is constructed, as in Table 5, where Latent Class 5 and a 'No' response are the reference pair and Latent Class 2 and a 'Yes' response are the comparison pair.

Table 5
The 2×2 table comparing Latent Classes 5 (Experimenters) and 2 (Regular smokers) on the ever smoke 100 or more cigarettes item. Quantities shown are the conditional response probabilities

Latent class	P (No|L)	P (Yes|L)
Have you smoked at least 100 cigarettes in your life?		
Experimenters	0.906	0.094
Regular smokers	0.019	0.981

Eq. (5) shows how to compute the odds ratio based on Table 5.

$$\hat{\psi} = \frac{(0.906)(0.981)}{(0.094)(0.019)} \approx 500 \quad (5)$$

Members of Latent Class 5 are approximately 500 times more likely to respond ‘No’, to Item 1 than members of Latent Class 2. This odds ratio is very large, confirming what the $\hat{\rho}$’s show = Latent Classes 2 and 5 reliably predict a ‘Yes’ and a ‘No’, respectively, to Item 1. Furthermore Yule’s Q transform of $\hat{\psi}$ in Eq. (5) results in a correlation greater than 0.99.

In contrast, consider the comparison of Latent Classes 3 and 4 on Item 2. In this case, $\hat{\psi} = 0.459(0.353 \times 0.457/0.543 \times 0.647)$, in other words, members of latent class 3 are about half as likely as members of latent class 4 to indicate ‘No’ to Item 2. The Yule’s Q of -0.37 shows that these two classes and item responses are not strongly correlated. From this, it appears that Item 2 is not a reliable indicator for these two classes.

When there are multiple response categories for a single item, more odds ratios are required to describe the possible 2×2 tables. As an example, consider the fourth item, ‘Do you think you will be smoking in 1 year?’ Since the equality constraints among the $\hat{\rho}'$s, latent classes 1 and 3 may be compared with latent classes 4 and 5 (see Table 6).

Eqs. (6)–(8) show the $\hat{\psi}$’s based on Table 6. In all cases, Classes 1 and 3 are more likely to endorse the lower response category (i.e. more negative beliefs about future cigarette smoking), than are members of Classes 4 and 5.

$$\hat{\psi}_{\mathrm{RC1vs.RC2}} = \frac{(0.756)(0.558)}{(0.227)(0.189)} \approx 11.90, \ \text{Yule's } Q = 0.84 \quad (6)$$

$$\hat{\psi}_{\mathrm{RC1vs.RC3}} = \frac{(0.756)(0.353)}{(0.017)(0.089)} \approx 179.71, \ \text{Yule's } Q = 0.99 \quad (7)$$

$$\hat{\psi}_{\mathrm{RC2vs.RC3}} = \frac{(0.227)(0.353)}{(0.017)(0.558)} \approx 8.45, \ \text{Yule's } Q = 0.79 \quad (8)$$

The last type of item-specific reliability is the probability of class membership conditional on item response (Eq. (3)). When working with continuous variables, reliability is typically conceived of as a correlation and therefore, the relation between items and factors is symmetric. However, with categorical data, relations are not necessarily symmetric. Therefore, examining the predicted latent class conditional on item response can provide a different picture of item reliability. Table 7 shows those estimates.

Where the $\hat{\rho}$’s sum to one across the response categories within a latent class for a single item, these estimates sum to one across the latent classes within a response category for a single item. Consider someone who said ‘No’ to the ever smoke 100+ item. If this is all that was known about this individual, according to this analysis, the chance that he or she is a member of Latent Class 1 is 0.82, whereas the chance that this person is in Latent Class 5 is only 0.13.

Consider the estimate for the regular smokers. Item 2, Response Category 2. This ρ was fixed to 1.0. However, someone who replied ‘Yes’ to the ever smoked every day for a month item has only a 0.68 chance of membership in the regular smoker class. Even though membership in Latent Class 2 perfectly predicts this item response, a ‘Yes’ response to this item does not perfectly predict membership in this class.

Even though the $\hat{\rho}$’s and odds ratios indicate that some items reliably measure each of the last three classes, this is not clear from Table 7. There are no item responses that singly predict membership in any of the last three latent classes above probability 0.4. The reason for this apparent discrepancy between reliability indices is that the estimates in Table 7 take the $\hat{\gamma}$’s into account. The last three classes are small proportions of the sample. Therefore, the probability of randomly selecting a member of one of those classes is small, even with the added information of knowing a single item response.

The least reliable indicator in Table 7 is Response Category 2 for the number of days smoked in past 30. This response is not clearly related to any single class. In

Table 6
Table of $\hat{\rho}$’s for computing odds ratios comparing Item 4 response categories for Latent Classes (1 and 3) to Latent Classes (4 and 5)

Latent classes	P (Definitely not|L)	P (Probably not|L)	P (Affirmative L)
Do you think you will be smoking cigarettes 1 year from now?			
1 and 3	0.756	0.227	0.017
4 and 5	0.089	0.558	0.353

Table 7
Latent class membership probabilities conditional on item response. 'NA' denotes not applicable

Latent class	Ever smoke 100+cigarettes	Ever smoke every day for a month	Days smoked in past 30	Plan to smoke in 1 year
		Response category 1		
	No	No	None	Def. Not
Past experimenters	0.82	0.78	0.88	0.83
Regular smokers	0.01	0.00	0.00	0.02
Past smokers	0.04	0.04	0.11	0.12
Irregular smokers	0.00	0.05	0.00	0.01
Experimenters	0.13	0.12	0.00	0.02
		Response category 2		
	Yes	Yes	Some	Poss. Not
Past experimenters	0.00	0.00	0.25	0.46
Regular smokers	0.61	0.68	0.07	0.15
Past smokers	0.17	0.18	0.08	0.07
Irregular smokers	0.20	0.10	0.23	0.13
Experimenters	0.03	0.03	0.37	0.20
		Response category 3		
			30	Yes
Past experimenters	NA	NA	0.00	0.05
Regular smoker	NA	NA	1.00	0.67
Past smokers	NA	NA	0.00	0.01
Irregular smokers	NA	NA	0.00	0.11
Experimenters	NA	NA	0.00	0.17

fact, it is almost equally predictive of Latent Classes 1, 4 and 5. However, recall that the $\hat{\rho}$ for Latent Class 1 and Response Category 2 of this item is 0.12, whereas for Latent Classes 4 and 5 the $\hat{\rho}$ is 0.98. If it were known that a person was a member of Latent Class 4 or 5, it would be quite probable that he or she would report some days smoked in the past 30. However, if the only information about a person were the item response 'I have smoked between 1 and 29 of the past 30 days', one could narrow the choice of class membership down to three classes but could not do much better than that.

3.2.2. Multiple items or scale reliability

Now, consider the problem of assembling a cigarette smoking scale in order to identify these classes. A primary purpose of such a scale is to distinguish the latent classes reliably. How well do these four indicators do? Table 8 shows the item-set reliabilities for these data.

The column on the left in Table 8 shows the response patterns. For example, the first response pattern correspond to replying 'No' to the first two items,'None' to the past 30 day item and 'Definitely not' to the smoke in 1 year item. The remaining columns on the right contain the latent class membership probabilities associated with each response pattern. An individual who responds negatively to all four items is nearly certain to be a member of Latent Class 1.

Scanning down the column for Latent Class 1, there are four response patterns that strongly (defined here as probability greater than 0.8) predict membership in this latent class. Those four patterns all contain 'No' responses to the first two items. However, different responses on the second two items differentially predict membership in Latent Class 1. For example, if an individual reports some days smoked in the past 30, and that he or she will definitely not be smoking in 1 year, the probability of membership in Latent Class 1 is 0.85. But, if instead he or she reports no smoking in the past 30 days, and affirmatively to the item about smoking in 1 year, the probability of membership in latent class 1 is 0.93. The probability of Class 1 membership drops to 0.04 for a person reporting some days smoked in the past 30 and affirmatively to the smoking in 1 year item.

Latent Class 2 is also predicted well by the model. In fact, due to the restrictions placed on the $\hat{\rho}$'s for Latent Class 2, membership in this class is predicted perfectly by six response patterns. These six response patterns all include the response indicating smoking on each of the past 30 days.

Near the bottom of Table 8 there are two response patterns that might be expected to predict membership in Latent Class 2. These patterns are (2, 2, 1, 3) and (2, 2, 2, 3). Both patterns predict membership in Latent Class 2 less than 60% of the time. The former is an unusual response pattern, in that, it represents a person reporting high or regular past use and no smoking in the past month, combined with the belief that he or she will be smoking in 1 year. Replying yes to both past use items is a good indicator of membership in Latent Class 2, but

Table 8
Item-set reliabilities for all four items

	Latent class				
Response pattern	Past exp.	Regular smk.	Past smk.	Irregular smk.	Exp.
1111	**0.99**		0.03		
2111			**0.99**	0.01	
1311			**1.00**		
2211			**0.99**		
1121	**0.849**		0.03		0.12
2121			0.49	0.40	0.10
1221		0.03	**0.81**	0.01	0.17
2221		0.11	0.64	0.21	0.01
1131	NA	NA	NA	NA	NA
2131	NA	NA	NA	NA	NA
1231		**1.00**			
2231		**1.00**			
1112	**0.98**		0.02		0.01
2112			**0.87**	0.10	0.03
1212			**0.97**	0.02	0.03
2212		0.03	**0.92**	0.05	
1122	0.25		0.01	0.01	0.74
2122			0.04	0.76	0.20
1322		0.01	0.17	0.05	0.77
2222		0.19	0.09	0.70	
1132	NA	NA	NA	NA	NA
2132	NA	NA	NA	NA	NA
1233		**1.00**			
2232		**1.00**			
1113	**0.92**		0.02		0.06
2113			0.45	0.44	0.11
1213		0.05	0.75	0.01	0.19
2213		0.54	0.31	0.14	
1123	0.01			0.02	**0.95**
2123			0.01	0.79	0.20
1223		0.08	0.02	0.05	**0.85**
2223		0.58	0.01	0.40	0.01
1133	NA	NA	NA	NA	NA
2133	NA	NA	NA	NA	NA
1233		**1.00**			
2233		**1.00**			

Values less than 0.01 are blank. Values greater than 0.8 are in bold, NA denotes 'not applicable;' some response patterns have 0.0 probability of occurring due to restrictions in the final model. Some tows do not sum to 1.0 due to rounding. Abbreviations: exp. = experimenter; smk. = smoker; irreg. = irregular.

the absence of daily smoking in the past 30 days does not fit with the overall profile of Class 2.

Latent Class 3 is also predicted well by the model, although there are some response patterns associated with moderate probabilities of class membership. There are seven response patterns that predict Class 3 with probability greater than 0.8. The distinctive aspect of these seven response patterns is the occurrence of a 'Yes' to one or both past cigarette smoking items in combination with none or some past 30 day use and negative beliefs about smoking in 1 year.

Latent Classes 4 and 5 have fewer predictive response patterns. No response patterns predict membership in Latent Class 4 greater than 0.8, but three are greater than 0.7. These patterns are (2, 1, 2, 2), (2, 2, 2, 2) and (2, 1, 2, 3). The distinctive features of these three patterns is the combination of a 'Yes' response to the smoke 100+ item, indicating smoking on some of the past 30 days, and not reporting that one will definitely not smoke in 1 year. Two response patterns predict latent class 5 membership at 0.8 or better. Those patterns are (1, 1, 2, 3) and (1, 2, 2, 3). The key features for these two response patterns are a 'No' to the 100+ item combined with some smoking in the past 30 days and an affirmative response to the smoke in 1 year item.

4. Conclusions

The latent class model was used to identify classes of adolescents reporting some cigarette smoking and to examine the reliability of the items used in the analysis.

Classes corresponding to two of the stages in the model of nicotine dependence onset were identified (experimenters and regular smokers); however, other classes were also required to fit the data. A class labeled past experimenters was 56% of the sample. Two other classes were past smokers (8%) and irregular smokers (6%).

From this reliability analysis, one is able to determine how well classes are measured, and how the items perform individually and as a set. In this case, reliability was variable. The two largest classes, past experimenters and current smokers, were the most reliably measured. The three smaller classes, were less reliably measured.

For example, the probability of membership in Latent Class 4 (irregular smokers) was not higher than 0.8 for any of the response patterns. It is also interesting to note that $\hat{\rho}$ for Item 2 was near 0.5, meaning that membership in the irregular smoker class was not predictive of item response. Thus, only three of the four items were useful in identifying this class, which could contribute to the lower reliability for this class. Measurement of this class could be improved by adding an indicator designed to identify it. The purpose of such an additional item would be to better identify a pattern of responding that indicates a significant amount of lifetime smoking (not just experimenting) with sporadic smoking (not daily or a regular smoker); the adolescent analogue of adult 'chippers' (Shiffman et al., 1994). Examples might be 'Do you typically smoke cigarettes every day?' or 'I usually only smoke on weekends or at parties. (True or False)'.

One class of primary interest is Latent Class 5 (experimenters). This class was measured relatively well; specifically the combination of no or low past use, some recent use, and equivocality about future smoking predicted membership in this class. However, this class was also relatively small, and this may pose some difficulties in evaluating measurement. First, the item-specific reliability index predicting class membership from a single item response is not very informative for small classes. A second issue is that this class is a relatively short-lived state in many people's life (US Department of Health and Human Services, 1994; Choi et al., 1997). Most likely, individuals move relatively quickly from experimentation to either past use (i.e. the individual has quit) or to some level of regular use. Therefore, it is likely that in any study, the proportion of people in the experimentation class is relatively small, which in turn provides less information for measurement evaluation. More items might not be the best way to improve the measurement of this class. In this case, a more targeted or adaptive sampling design could be advantageous (see Thompson and Collins, 2002, in this issue).

4.1. Alternate approaches

There has been other work on the reliability of categorical data. For example, Collins (2001) presents a reliability index based on the latent class model that is designed to be similar to a reliability coefficient like Cronbach's α (Cronbach, 1951). An advantage of this approach is that it is a single value that is on a scale familiar to many researchers. However, it was not presented here because the current formulation does not allow missing data. Also, any single number necessarily glosses over differences in how reliably the classes are measured. For example, a single number cannot simultaneously indicate that measurement was very reliable for regular smokers, but not for irregular smokers.

Work in survey statistics has focused on reliability in categorical variables such as employment status and age group. In this literature, reliability is referred to as the simple response variance (Hansen et al., 1964; O'Muircheartaigh, 1991). However, these methods are more similar to test–retest methods, rather than multiple indicator models. Recently though, the latent class model is discussed in the statistics literature as a measurement error model (Biemer and Wiesen, 2002).

Kraemer (1992) presents a method for estimating the reliability of categorical variables in the context of medical research. These procedures are designed to evaluate the reliability of classifications made by multiple raters = for example, multiple doctors diagnosing patients. The classification being made (i.e. the categorical variable) is the class of interest, not one characteristic that distinguishes latent classes. As such, the classes need to be known before the measurement process can be designed.

4.2. Limitations

The data are one concern in this study. Specifically, there was less evidence of random error than would be expected. As is commonly done in survey research, the TAPS data were cleaned and edited prior to the release of public use data files (personal communication from Karen F. Allen, e-mail, 29 August 2001). Since the data cleaning procedures can potentially introduce bias, unverifiable response inconsistencies should be left unedited and modeled in the analysis when possible. This article has shown that random error in categorical data is easily handled by the latent class model. The ideal situation for public use data files like TAPS might be to distribute both edited and unedited versions of the data.

Another concern about this work might be thought of as over-fitting and the replicability of the latent class structure and the parameter restrictions. Specifically, parameter restrictions were applied sequentially leading to the final model. If a certain restriction was not acceptable, it was rejected. One could argue that this style of model fitting could result in overly good model fit. However, it is important to note that model complexity was sequentially reduced, not interested to suit the sample at hand. One could have stopped the model fitting with the unrestricted model. By imposing parameter restrictions, the model is made more parsimonious. An advantage of the use of parameter restrictions in a measurement study is that class similarities and differences are highlighted. For example, the past experimenters and past smokers reported similar beliefs about future smoking, but quite different patterns of past cigarette smoking. This further illustrates the response pattern orientation of latent class analysis. Believing that smoking in the next year is unlikely distinguishes past experimenters and past smokers from the other three classes. However, it is past cigarette smoking behavior that distinguishes the past experimenters from the past smokers. Substantively, it is interesting that both past experimenters and past smokers are equally disposed toward future smoking. The applicability of the parameter restrictions in this model can be tested in other samples.

This article has briefly introduced the latent class model, in order to focus on reliability assessment. Researchers interested in fitting these models are encouraged to read more detailed presentations (Goodman, 1974; McCutcheon, 1987; Heinen, 1996). Software to fit the latent class model and other categorical variable models is available. Two free programs are ℓEM (Vermunt, 1997) and WINLTA (Collins et al., 1999). Both are available on-line and both provide informative manuals and examples.

In conclusion, the latent class model has been shown to provide a rich picture of class structure within a data set and to assess the reliability of categorical items. Given that many substance use survey items are categorical in nature, these procedures may prove helpful in a variety of drug abuse research settings.

Acknowledgements

The author would like to thank the reviewers for very helpful comments. Due to their time and thoughtfulness, this article is much improved. This work has been supported by The Robert Wood Johnson Foundation Research Network on the Etiology of Tobacco Dependence (TERN) and by National Institute on Drug Abuse grant P50DA10075.

References

Agresti, A., 1990. Categorical Data Analysis. Wiley, New York.

Aitkin, M., Rubin, D.B., 1985. Estimation and hypothesis testing in finite mixture models. Journal of the Royal Statistical Society Series B 47, 67–75.

Allen, K.F., Moss, A.J., Giovino, G.A., Shopland, D.R., Pierce, J.P., 1992. Teenage tobacco use: data estimates from the Teenage Attitudes and Practices Survey. United States, 1989. Advance Data From Vital and Health Statistics: no. 224, National Center for Health Statistics, Hyattsville, MD.

Bartholomew, D.J., 1987. Latent Variable Models and Factor Analysis. Oxford University Press, New York.

Biemer, P.B., Wiesen, C., 2002. Latent class analysis of embedded repeated measurements: an application to the national household survey on drug abuse. Journal of the Royal Statistical Society, Series A 165, 97–119.

Carmines, E.G., Zeller, R.A., 1979. Reliability and Validity Assessment. Sage Publications, Newbury Park, CA.

Centers for Disease Control and Prevention, 1991. Cigarette Smoking Among Youth—United States, 1989. Morbidity and Mortality Weekly Report 40, pp. 712–715.

Choi, W.S., Pierce, J.P., Gilpin, E.A., Farkas, A.J., Berry, C.C., 1997. Which adolescent experimenters progress to established smoking in the United States. American Journal of Preventive Medicine 13, 385–391.

Clogg, C.C., 1995. Latent class models. In: Arminger, G., Clogg, C., Sobel, M. (Eds.), Handbook of Statistical Modeling for the Social and Behavioral Sciences. Plenum Press, New York, pp. 311–359.

Clogg, C.C., Manning, W.D., 1996. Assessing reliability of categorical measurements using latent class models. In: von Eye, A., Clogg, C. (Eds.), Categorical Variables in Developmental Research. Academic Press, San Diego, CA, pp. 169–182.

Collins, L.M., 2001. Reliability for static and dynamic categorical latent variables: Developing measurement instruments based on a model of the growth process. In: Collins, L., Sayer, A. (Eds.), New Methods for the Analysis of Change. American Psychological Association, Washington, DC, pp. 273–288.

Collins, L.M., Flaherty, B.P., Hyatt, S.L., Schafer, J.L., 1999. WINLTA User's Guide 2.0. The Pennsylvania State University, University Park, PA.

Cook, T.D., Campbell, D.T., 1979. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Houghton Mifflin, Boston, MA.

Cox, B.G., Witt, M.B., Traccarella, M.A., Perez-Michael, A.M., 1992. Inconsistent reporting of drug use in 1988. In: Turner, C., Lessler, J., Gfroerer, J. (Eds.), Survey Measurement of Drug Use: Methodological Studies DHHS Publication No. (ADM) 92-1929). US Department of Health and Human Services, Washington, DC.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334.

Darling, N., Cumsille, P., 2002. Theory, measurement, and methods in the study of family influences on adolescent smoking. Addiction, in press.

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

Flay, B.R., 1993. Youth tobacco use: risks, patterns, and control. In: Orleans, C., Slade, J. (Eds.), Nicotine Addiction. Oxford University Press, New York, pp. 360–384.

Golub, A.L., Johnson, B.D., 2001. Variation in youthful risks of progression from alcohol and tobacco to marijuana and to hard drugs across generations. American Journal of Public Health 91, 225–232.

Golub, A., Labouvie, E., Johnson, B.D., 2000. Response reliability and the study of adolescent substance use progression. Journal of Drug Issues 30, 103–108.

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

Gorsuch, R.L., 1983. Factor Analysis, second ed. Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

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

Guilford, J.P., 1954. Psychometric Methods, second ed.. McGraw-Hill, New York.

Guo, J., Collins, L.M., Hill, K.G., Hawkins, J.D., 2000. Development pathways to alcohol abuse and dependence in young adulthood. Journal of Studies on Alcohol 61, 799–808.

Hansen, M.H., Hurwitz, W.N., Pritzker, L., 1964. The estimation and interpretation of gross differences and the simple response variance. In: Rao, C.R. (Ed.), Contributions to Statistics. Pergamon Press, Oxford, pp. 111–136.

Harrison, L., Hughes, A. (Eds.), 1997. The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates. NIDA Research Monograph 167. Department of Health and Human Services, National Institute on Drug Abuse, Division of Epidimiology and Preventive Research, Rockville, MD.

Heinen, T., 1996. Latent Class and Discrete Latent Trait Models. Sage Publications, Thousand Oaks, CA.

Kandel, D.B., Kessler, R.C., Margulies, R.Z., 1987. Antecedents of adolescent initiation in stages of drug use: a developmental analysis. Journal of Youth and Adolescence 7, 646–676.

Kandel, D.B., Yamaguchi, K., Chen, K., 1992. Stages of progression in drug involvement from adolescence to adulthood: further evidence for the gateway theory. Journal of Studies on Alcohol 53, 447–457.

Kane, R.J., Yacoubian, G.S., Jr, 1999. Patterns of drug escalation among Philadelphia arrestees: an assessment of the gateway theory. Journal of Drug Issues 29, 107–120.

Kerlinger, F.N., 1964. Foundations of Behavioral Research. Holt, Rinehart and Winston, New York.

Kraemer, H.C., 1992. Measurement of reliability for categorical data in medical research. Statistical Methods in Medical Research 1, 183–199.

Kuder, G.F., Richardson, M.W., 1937. The theory of the estimation of test reliability. Psychometrika 2, 151–160.

Leventhal, H., Cleary, P.D., 1980. The smoking problem: a review of the research and theory in behavioral risk modification. Psychological Bulletin 88, 370–405.

Lord, F.M., Novick, M.R., 1968. Statistical Theories of Mental Test Scores. Addison-Wesley, Reading, MA.

Maisto, S.A., McKay, J.R., Connors, G.J., 1990. Self-report issues in substance abuse: State of the art and future directions. Behavioral Assessment 12, 117–134.

Massey, J.T., Moore, T.F., Parsons, V.L., Tadros, W., 1989. Design and estimation for the national health interview survey, 1985–1994. Vital Health Statistics 2, Number 110, National Center for Health Statistics, Hyattsville, MD.

Mayhew, K.P., Flay, B.R., Mott, J.A., 2000. Stages in the development of adolescent smoking. Drug and Alcohol Dependence 59, S61–S81.

McCutcheon, A.L., 1987. Latent Class Analysis. Sage Publications, Thousand Oaks, CA.

Moss, A.J., Allen, K.F., Giovino, G.A., Mills, S.L., 1992. Recent trends in adolescent smoking, smoking-uptake correlates, and expectations about the future. Advance Data from Vital and Health Statistics; No. 221. National Center for Health Statistics, Hyattsville, MD.

O'Muircheartaigh, C., 1991. Simple response variance: estimation and determinants. In: Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., Sudman, S. (Eds.), Measurement Errors in Surveys. Wiley, New York, NY, pp. 551–574.

Peters, D., 1997. A natural classification of alcoholics by means of statistical grouping methods. Addiction 92, 1649–1661.

Raftery, A.E., 1995. Bayesian model selection in social research (with discussion). In: Marsden, P. (Ed.), Sociological Methodology. Blackwells, Cambridge, MA, pp. 111–196.

Sakamoto, Y., 1991. Categorical Data Analysis by AIC. Kluwer Academic Publishers, Tokyo, Japan.

Schafer, J.L., 1997. Analysis of Incomplete Multivariate Data. Chapman Hall, London.

Shiffman, S., Kassel, J.D., Paty, J.A., Gnys, M., 1994. Smoking typology profiles of chippers and regular smokers. Journal of Substance Abuse 6, 21–35.

Spearman, C., 1910. Correlation calculated with faulty data. British Journal of Psychology 3, 271–295.

Thompson, S.K., Collins, L.M., 2002. Adaptive sampling in research on risk-related behaviors. Drug and Alcohol Dependence 68 (Suppl. 1) 57–68.

Turner, C.F., Lessler, J.T., Gfroerer, J.C. (Eds.), 1992. Survey Measurement of Drug Use: Methodological Studies DHHS Publication No. (ADM) 92-1929 US Department of Health and Human Services, Washington, DC.

US Department of Health and Human Services, 1994. Preventing Tobacco Use Among Young People: A Report of the Surgeon General. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health, Atlanta, GA, US.

Vermunt, J.K., 1997. ℓEM: a General Program for the Analysis of Categorical Data. Tilburg University.