3

A PARTIAL INDEPENDENCE ITEM RESPONSE MODEL FOR SURVEYS WITH FILTER QUESTIONS

Sean F. Reardon*
Stephen W. Raudenbush†

In many surveys, responses to earlier questions determine whether later questions are asked. The probability of an affirmative response to a given item is therefore nonzero only if the participant responded affirmatively to some set of logically prior items, known as "filter items." In such surveys, the usual conditional independence assumption of standard item response models fails. A weaker "partial independence" assumption may hold, however, if an individual's responses to different items are independent conditional on the item parameters, the individual's latent trait, and the participant's affirmative responses to each of a set of filter items. In this paper, we propose an item response model for such "partially independent" item response data. We model such item response patterns as a function of a person-specific latent trait and a set of item parameters. Our model can be seen as a generalized hybrid of a discrete-time hazard model and a Rasch model. The proposed procedure yields estimates of (1) person-specific, interval-scale measures of a latent trait (or traits), along with person-specific standard errors of measurement; (2) conditional and marginal

*Stanford University
†University of Chicago

item severities for each item in a protocol; (3) person-specific conditional and marginal probabilities of an affirmative response to each item in a protocol; and (4) item information and total survey information. In addition, we show here how to investigate and test alternative conceptions of the dimensionality of the latent trait(s) being measured. Finally, we compare our procedure with a simpler alternative approach to summarizing data of this type.

1. INTRODUCTION

In social surveys, it is common to inquire about whether events have occurred, and if so, to inquire about specific aspects of these events, such as their frequency and intensity. Examples include substance use (whether one has used a substance, and if so how often and how heavily); crime (whether one has been involved in a given type of crime, and if so, how often); symptoms of a disease; purchases of a given product. Protocols of this type have a conditional structure: responses to earlier questions determine whether later questions are asked. The researcher presumes that the entire ensemble of responses carries useful information about one or more underlying "latent" behavioral traits or attributes.

Unfortunately, standard latent variable models for item response data are not suited for the analysis of data having a conditional structure. Instead, such models assume local independence—that is, conditional independence of all item responses given the item parameters and the person traits being measured (Lord & Novick 1968). Such an assumption cannot hold when responses to a prior item determine whether a later item is asked. However, a weaker "partial independence" assumption may hold, where by "partial independence" we mean local independence of item responses given a participant's affirmative responses to each of a set of logically prior items.

In this paper, we propose a principled procedure for modeling response patterns to such protocols as a function of a person-specific latent trait and a set of item parameters. Our procedure yields estimates of four quantities of interest: (1) person-specific, interval-scale measures of a latent trait (or traits), along with person-specific standard errors of measurement; (2) conditional and marginal item severities for each item in a protocol; (3) person-specific conditional and marginal probabilities

of an affirmative response to each item in a protocol; and (4) item information and total survey information. In addition, we show here how to investigate and test alternative conceptions of the dimensionality of the latent trait(s) being measured. Finally, we compare our procedure with a simpler alternative approach to summarizing data of this type.

We achieve these aims by formulating fixed and random effects models for item responses that may be conditional on prior responses while being independent of responses to other items. The modeling approach can be seen as a generalized hybrid of a discrete-time hazard model and a Rasch model. The conditional structure of the items enables us to define "risk sets" of individuals who possess a nonzero probability of responding affirmatively to sets of logically subsequent items; the logic of the discrete-time hazard model (Allison 1982) and the (formally equivalent) continuation ratio model for ordinal outcomes (Armstrong and Sloan 1989; Cox 1988; Fienberg 1980) provides guidance here. Unlike the discrete-time and continuation ratio models, however, our model incorporates the possibility that multiple "locally-independent" items may be asked of each individual in a risk set. The model assumes that each person possesses a latent trait or attribute that affects the probability of an affirmative response to each such locally independent item, conditional on having affirmatively answered all logically prior items. In addition, each such item may have subsequent items conditional upon it, allowing us to fit models based on complex sets of both conditional and independent items.

Although the approach we develop in this paper has application to a wide range of substantive survey domains, we focus on a single potential application—the problem of estimating latent alcohol and marijuana use from a set of survey items regarding substance use in the last year—in order to provide a concrete illustration of the method. In particular, we demonstrate the application of our model to study the alcohol and marijuana use of a large, diverse, and representative sample of children growing up in Chicago. To illustrate how our model can be used to investigate the dimensionality of latent characteristics, we then investigate whether adolescent alcohol and marijuana use can be considered aspects of a single underlying latent substance use trait, or whether they represent distinct behavioral traits.

2. BACKGROUND AND SIGNIFICANCE

Researchers interested in adolescent substance abuse—as well as in a wide range of other substantive domains—have contended with the conditional structure of survey data in a variety of ways. Each has some utility in particular circumstances, but none of the available methods used to date efficiently combines information across all item responses within the framework of latent trait analysis, which is the aim of the current paper.

2.1. *Single-Item Analysis*

A simple strategy for analyzing data having a conditional structure is to study one item at a time (Adalbjarnardottir 2002; Bailey, Flewelling, and Rachal 1992; Chassin, Pitts, and Prost 2002; Hill et al. 2000; Khoo and Muthen 2000). Suppose, for example, that the first question on a survey is "Have you had a drink during the past year?" Modeling responses to such a question as a function of covariates or exposure to prevention programs could be quite useful.

One obvious limitation of single-item analysis is that it does not allow a pooling of information across item responses to reduce the error with which a latent trait is measured. This limitation of single-item analysis provides an important rationale for latent trait models for multiple item responses. Latent trait models also supply a basis for quantifying measurement error, a benefit that single-item analysis cannot enjoy. Moreover, if a questionnaire includes many items, single-item analysis applied to each item will give rise to a large number of hypothesis tests, increasing the risk of a Type I error. It may prove difficult to summarize evidence across the many analyses so generated.

The limitations of single-item analysis described above apply to all multiple-item surveys, even those without a conditional structure. However, when single-item analysis is applied to survey items having a conditional structure, a subtle and potentially pernicious additional concern arises: conditioning on an error-prone response. To illustrate, consider a case in which persons who respond affirmatively to the question "Have you had a drink in the past year?" are asked "Were you drunk in the last year?" Applying single-item analysis to this second, conditional question requires that those who responded negatively to the first

question be discarded. Yet, some of those cases may be discarded simply as a result of measurement error.

Consider a thought experiment in which two persons with the same propensity to drink respond differently to the first question because of differences in estimating when they last had a drink or because the timing of the last drink differed slightly. For example, one person might have had a drink 364 days ago and correctly responded "yes" to the first question while a second person with an identical propensity to drink may have had a drink 366 days ago and thus correctly responded "no" to the first question. In this case, single-item analysis for the second item would require that the second person be discarded, strictly as a result of measurement error. Selecting the sample for an analysis conditional on measurement error may then create unwanted results such as regression to the mean. The only way to avoid such a problem using single-item analysis is to restrict application to nonconditional items. But such a procedure then requires that information from all conditional items be discarded, meaning that the expense required to collect this additional information will have been wasted.

2.2. *Multiple Item Analysis*

A second strategy is to develop a transformation that combines multiple-item responses to a single variable believed to determine these multiple-item responses. This strategy is widely used. (Recent examples are found in Barnes et al. 2000; Bennett et al. 1999; Colder and Stice 1998; Duncan, Duncan, and Hops 1996; Hussong, Curran, and Chassin 1998; Scheier et al. 2000; Schulenberg et al. 1996; Silberg et al. 2003; Wills and Cleary 1999.) In this approach, item responses might be added, for example, or combined into ordered categories. In comparison to single-item analysis, such an approach has the advantage of combining more information from the survey to measure the variable of interest, in principle reducing measurement error and also allowing a more parsimonious analytic plan. A limitation of this method, however, is that it is not based on a probabilistic model for how the item responses are generated. Thus, the benefits of latent trait analysis via item response modeling are not available. These include a principled calibration of items, study of item fit, the quantification of measurement error, and the evaluation of dimensionality.

The study of dimensionality has emerged as an important topic in item response modeling (Adams, Wilson, and Wang 1997; Cheong and Raudenbush 2000; Raudenbush, Johnson, and Sampson 2003; Reckase 1985). Study of dimensionality has important implications for the assessment of construct validity. Suppose, for example, that responses to items about alcohol use and marijuana use are treated as a single dimension when in truth two separate dimensions, one for each substance, are driving the item responses. Such a unidimensional analysis would fail to reveal the different processes that predict use of the two substances. Multidimensionality takes a different form when subgroups (e.g., males and females) respond differently to particular items even holding constant the latent trait of interest. Such differential item functioning (DIF; Holland & Wainer 1993), sometimes called "item bias," can be studied in a principled way using item response modeling, but this benefit has not yet been extended to surveys having a conditional item structure.

Finally, when the items have a partially conditional structure, constructing a variable by a simple transformation of multiple-item responses may propagate measurement error. Let us again take up the example described above, in which two persons having the same propensity to use alcohol responded differently to the question "Have you used alcohol during the past year?" In this case, the person responding affirmatively, but not the person responding negatively, will be asked more questions, perhaps many more questions, increasing the opportunity for the two cases to be incorrectly differentiated when all item responses are combined.

2.3. *The Case for a Latent Trait Model for Partially Independent Item Response Data*

The foregoing discussion suggests that the advantages afforded by latent trait analysis ought to be extended to survey data having a partially conditional structure. This requires a reasonable probabilistic model to describe responses to a mixture of item types. We shall define a "gate" item as one that must be answered affirmatively if a set of logically subsequent items is to be administered to a given respondent. Individuals who answer a given "gate" item or set of items affirmatively are then in the "risk set" for logically subsequent items. Our probabilistic approach

regards the marginal probability of an affirmative response to a given item as the product of the probability of responding affirmatively to its gate item(s) (i.e., being in the risk set for the given item) and the conditional probability of responding affirmatively to that item given that one responded affirmatively to its gate item(s). Our approach thus combines a hazard model of the probability of passing a gate, given that prior gates have been passed, and a Rasch model to describe variation in the probabilities of responding affirmatively to sets of items, conditional on being in the risk set for each item. To illustrate, we apply this approach to a large and representative sample of data on children growing up in Chicago.

3. SAMPLE AND DATA

For the illustrative examples in this paper, we use data on adolescent substance use from a subsample of the longitudinal cohort study of the Project on Human Development in Chicago Neighborhoods (PHDCN), an ongoing, multilevel, prospective, longitudinal study designed to investigate the effects of neighborhood demographic and social context on a wide array of developmental and behavioral outcomes. The PHDCN study consists of a representative sample of children and youth living in 80 neighborhoods of Chicago between 1995 and 1996. The sampling design and data collection procedures for PHDCN are described in detail in Sampson, Morenoff, and Raudenbush (2005).

Our sample consists of adolescents from the 12-, 15-, and 18-year-old cohorts of the PHDCN sample. Each adolescent was interviewed three times, at roughly two- to three-year intervals. For each adolescent in our subsample, we use only one of three available longitudinal observations for the analyses reported here—wave 1 interview data from the 18-year-old cohort; wave 2 interview data from the 15-year-old cohort; and wave 3 interview data from the 12-year-old cohort. We restrict the age range in our sample in order to avoid confounding measurement issues with potential age variation in the measurement model. Of the 1531 eligible subjects, we exclude 79 subjects (5%) missing data on any of the four alcohol items (70 subjects) or on any of four demographic variables—age, gender, race/ethnicity, and socioeconomic status (9 subjects). Thus our subsample consists of interview data from a

representative sample of 1452 Chicago adolescents, aged 16–19 between 1995 and 2001 (mean age=17.4, S.D.=.8; 53% female; 19% white, 37% black; 44% Latino).

Adolescents in the PHDCN sample were asked a set of questions about their alcohol use in the last year. We use information from four alcohol use questions:

1. *How many times did you drink alcohol in the last year?* (possible responses were: never, 1–2 times, 3–5 times, 6–11 times, 12–24 times, 25–50 times, 51–99 times, 100–199 times, 200 or more times).
2. *How many times were you drunk in the last year?* (asked of those whose response to question 1 indicated they had a drink at least once in the last year—subjects could answer with any number; responses ranged from 0 to 300).
3. *How many times did you drink in the last month?* (asked of those whose response to question 1 indicated they had a drink at least once in the last year—possible responses were: never, 1–2 times, 3-5 times, 6–9 times, 10–14 times, 15–20 times, 21 or more times).
4. *How many times did you have more than 5 drinks in a row in the last month?* (asked of those whose response to question 3 indicated they had a drink at least once in the last month—possible responses were: never, 1 time, 2 times, 3–5 times, 6–9 times, 10 or more times).

From these four questions, we construct 11 binary items, each of which indicates whether an individual's response to a specific question is at or above a certain threshold level. For example, using thresholds of 1, 6, and 25 for the first question, we construct three items: *Did you drink at least 1 time in the last year*? *Did you drink at least 6 times in the last year*? *Did you drink at least 25 times in the last year*? For question 2 we use thresholds of 1, 6, and 25; for question 3, we use thresholds of 1, 3 and 10; and for question 4 we use thresholds of 1 and 3.

This procedure results in 11 binary items (see Table 1). Each of these items has a "risk set"—the set of persons who could have logically answered '*yes*' to the item, given their prior responses. This is the set of persons who answered *yes* to the logically prior item or items (e.g., the risk set for the '*had at least one drink in the last month*' item is the set of persons who said *yes* to the '*had at least one drink in the last year*' item). The conditional probability for an item is the probability of saying yes to the item, conditional on being in the risk set for the item.

TABLE 1
Observed Alcohol Item Response Frequencies

Item	Description	Gate Items	Marginal Probability	Risk Set	Number of *yes*	Conditional Probability
1	Alcohol in year	-none-	.554	1,452	805	.554
2	Alcohol 6x in year	1	.262	805	380	.472
3	Alcohol 25x in year	1, 2	.116	380	169	.445
4	Drunk in year	1	.341	805	495	.615
5	Drunk 6x in year	1, 4	.094	495	137	.277
6	Drunk 25x in year	1, 4, 5	.038	137	55	.401
7	Alcohol in month	1	.320	805	464	.576
8	Alcohol 3x in month	1, 7	.158	464	229	.494
9	Alcohol 10x in month	1, 7, 8	.039	229	57	.249
10	Binge in month	1, 7	.160	464	233	.502
11	Binge 3x in month	1, 7, 10	.064	233	93	.399

The marginal probability for an item is simply the probability of saying *yes* to the item.

Although we could have constructed more items or fewer items by selecting a different set of thresholds, our choice of thresholds was not arbitrary. We chose thresholds that would differentiate well across the range of responses and would give conditional probabilities far from 0 or 1, since constructing the items this way preserves most of the information in the responses while keeping the number of constructed items relatively parsimonious.[1]

4. THE MODEL

4.1. *Item Structure and the Gate Matrix*

Suppose that persons $i = 1,\ldots,n$ respond to items $k = 1,\ldots,K$, generating, for each person, the response vector $\mathbf{Y}_i = [y_{i1}, y_{i2}, \ldots, y_{iK}]$, where

---

[1] In additional analyses (not shown) we use 8 thresholds for questions 1 and 2 (1, 3, 6, 12, 25, 51, 100, 200), 6 thresholds for question 3 (1, 3, 6, 10, 15, 21), and 5 thresholds for question 4 (1, 2, 3, 6, 10). This uses the maximum possible information from questions 1, 3, and 4, and most of the information from question 2, and results in 27 items rather than 11. Results based on this more detailed set of conditional items are not substantially different from those based on the more parsimonious set of items, so we present the more parsimonious set here for simplicity of presentation.

$y_{ik} = 1$ if person i responded affirmatively to item k, and $y_{ik} = 0$ if person i responded negatively to item k or if person i did not respond to item k because he or she responded negatively to a gate item for item k, and so was not asked item k. We indicate the item structure as follows: we say that item j is a gate for item k if $y_j = 0$ implies that $y_k = 0$. We define the $K \times K$ gate matrix G, where element $G[k,j] = G_{kj} = 1$ if item j is a gate item for item k, and 0 otherwise (by definition, $G_{kk} = 0$, since an item cannot be a gate item for itself). The pattern of 1s in a given row k of G indicates which items are gate items for item k; the pattern of 1s in a given column j of G indicates which items are conditional on item j (items for which j is a gate item).

Next we define h_{ik}, the gate value for item k for person i, as follows: define $h_{ik} = 1$ if $y_{ij} = 1$ for all items j that are gate items for k, and $h_{ik} = 0$ if $y_{ij} = 0$ for at least one item j that is a gate item for k:

$$h_{ik} = \prod_{j=1}^{K} \left[1 - G_{kj} \left(1 - y_{ij} \right) \right]. \tag{1}$$

In other words, h_{ik} indicates whether the gate item or items for item k for person i are satisfied, and so indicates whether person i is in the risk set for item k. In a survey, h_{ik} will be 0 for items that person i is not asked because he or she did not meet the conditions necessary to be asked them (because his or her response to item k is determined by his or her negative response(s) to one or more of the gate items for item k). We denote the risk set R_k for item k as the set of all individuals i who have $h_{ik} = 1$.
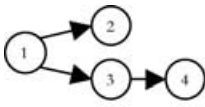
To make this more concrete, consider a simple set of three items, where all individuals are asked item 1, but where items 2 and 3 are asked conditional on an affirmative response to item 1. Such an item structure can be illustrated as follows:
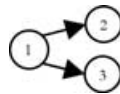
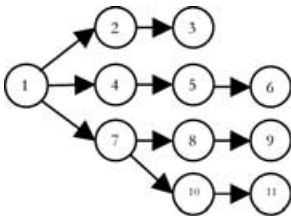For this item structure, the gate matrix G would be

$$G = \begin{bmatrix} 0\,0\,0 \\ 1\,0\,0 \\ 1\,0\,0 \end{bmatrix}. \qquad (2)$$

If we added a fourth item that was conditional on item 3, the item structure and corresponding gate matrix would be

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}. \qquad (3)$$

For our alcohol data, the item structure and corresponding gate matrix defined by the questionnaire skip fields and the thresholds we use to define the binary items are (where the items are numbered as in Table 1):

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} \qquad (4)$$

As an aside, note that the gate matrix can be more complex than this: In our case, item 2 (*drank alcohol 6 or more times in the last year*), for example, is a logical gate for item 5 (*was drunk 6 or more times in the last*

year), even though this is not reflected in the skip patterns of the survey questionnaire. If we incorporated this logical gate (and other similar ones) into the gate matrix, we would have

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix} \tag{5}$$
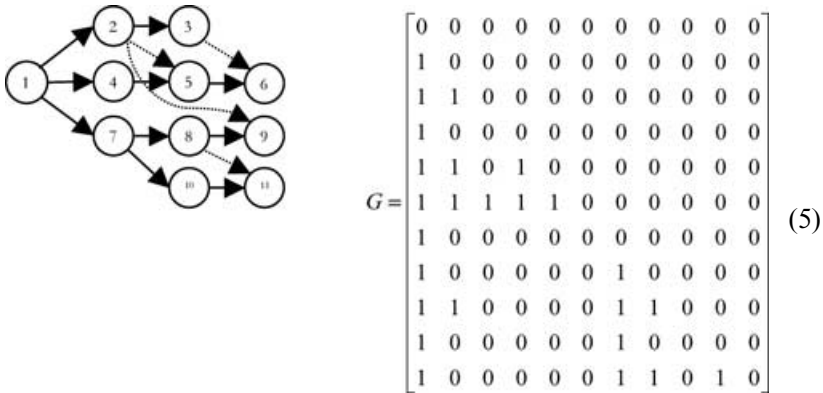
In our illustrative example for this paper, we use gate matrix (4) above, though the model applies equally well to a matrix like that in (5).[2]

4.2. *Model Notation*

Let ϕ_{ik} denote the conditional probability of a *yes* response to item k for person i (conditional on a *yes* response to all gate items for item k). That is,

$$\phi_{ik} = \Pr(y_{ik} = 1 | h_{ik} = 1). \tag{6}$$

---

[2]The problem with (5) is that the data may not exactly correspond to the logical gate matrix in (5): for example, some subjects may have said they were drunk 6 or more times in the last year but also said that they had not had a drink 6 or more times in the last year. To use the logical gate matrix in (5) with such data requires us to decide that one of these two responses is incorrect and to recode it. To use the questionnaire gate matrix in (4) allows us to use the responses as provided (but erroneously assumes the responses to the two items in question are conditionally independent, which is incorrect—item 5 is conditional on item 2). In our data, there are 22 cases (1.5% of 1452 total cases) containing logically inconsistent responses to the four alcohol use questions (e.g., a subject reported drinking fewer than 6 times in the last year, but reported being drunk 6 or more times in the last year, or drinking 10 or more times in the last month).

We assume that each individual has a conditional probability of saying *yes* to each item k, even if he or she is not asked item k because he or she does not say *yes* to the necessary gate item(s). This conditional probability will be a function of θ_i, the unobserved latent trait for person i. In the fixed effects models below, θ_i is a fixed parameter, while in the random effects models, we shall assume $\theta_i \sim N(0, \tau)$.

Finally, let π_{ik} denote the marginal probability that person i responds affirmatively to item k:

$$\begin{aligned} \pi_{ik} &= \Pr(y_{ik} = 1) \\ &= \phi_{ik} \cdot \pi_{iR_k}, \end{aligned} \tag{7}$$

where π_{iR_k} is the marginal probability that person i is in the risk set for item k:

$$\begin{aligned} \pi_{iR_k} &= \Pr(h_{ik} = 1) \\ &= \prod_{j=1}^{K} \phi_{ij}^{G_{kj}}. \end{aligned} \tag{8}$$

Thus we have

$$\begin{aligned} \pi_{ik} &= \phi_{ik} \cdot \pi_{iR_k} \\ &= \phi_{ik} \prod_{j=1}^{K} \phi_{ij}^{G_{kj}} \end{aligned} \tag{9}$$

To summarize our notation:

k indexes items.

i indexes persons.

y_{ik} is the observed response to item k for person i.

Y_i is the observed response vector for person i.

h_{ik} is the observed gate value for item k for person i.

H_i is the observed vector of gate values for person i.

R_k is the risk set for item k.

G_{kj} indicates whether item k is conditional on item j.

ϕ_{ik} denotes the conditional probability of a *yes* response to item k for person i.

π_{ik} denotes the marginal probability of a *yes* response to item k for person i.

θ_i denotes the (unobserved) latent trait for person i.

γ_k denotes the conditional severity of item k.

τ denotes the variance of θ_i.

4.3. *Fixed Effects Likelihood*

We model the conditional probability of a positive response to item k as

$$\phi_{ik} = \left[1 + e^{-(\theta_i - \gamma_k)}\right]^{-1}. \tag{10}$$

Under the fixed effects specification, the parameters are $\gamma_1, \ldots, \gamma_K; \theta_1,$ $\ldots, \theta_n$. The probability of observing response pattern $\mathbf{Y}_i$ for person i is then given by

$$\Pr(\mathbf{Y}_i) = \prod_{k=1}^{K} \left[\phi_{ik}^{y_{ik}} (1 - \phi_{ik})^{(1 - y_{ik})}\right]^{h_{ik}} \tag{11}$$

From this, we write the fixed effects log-likelihood, l_f, of observing the response pattern $\mathbf{Y}$ found in the data, where $\gamma = (\gamma_1, \ldots, \gamma_K)$ and $\theta = (\theta_1, \ldots, \theta_n)$, as

$$
\begin{aligned}
l_f = \ln[L(\mathbf{Y}; \gamma, \theta)] &= \sum_{i=1}^{n} \sum_{k=1}^{K} h_{ik} \left[y_{ik} \ln\left(\tfrac{\phi_{ik}}{1 - \phi_{ik}}\right) + \ln(1 - \phi_{ik})\right] \\
&= \sum_{k=1}^{K} \sum_{i \in R_k} [y_{ik} \eta_{ik} + \ln(1 - \phi_{ik})] \\
&= \sum_{k=1}^{K} l_{fk} \\
&= \sum_{i=1}^{n} l_{fi},
\end{aligned}
\tag{12}
$$

where $l_{fk} = \sum_{i=1}^{n} h_{ik} [y_{ik} \eta_{ik} + \ln(1 - \phi_{ik})]$ is the fixed effects log-likelihood for item k, given γ_k and θ; $l_{fi} = \sum_{k=1}^{K} h_{ik} [y_{ik} \eta_{ik} + \ln(1 - \phi_{ik})]$ is the fixed effects log-likelihood for person i, given γ and θ_i; and $\eta_{ik} = \ln[\phi_{ik}/(1 - \phi_{ik})]$. Note that $y_{ik} = 0$ by definition when $h_{ik} = 0$, so l_{fk} is the log-likelihood of observing the pattern of responses to item k for all who satisfied the gate items to item k (all those in R_k). The observed (structural) 0s for item k among those who did not satisfy the gate items for k contribute no information

to the log-likelihood l_{fk} since they are determined by the condition $h_{ik} = 0$.

Since l_{fk} depends only on those who can possibly have answered item k, we can drop item k for person i from the data set if $h_{ik}=0$. This is what we do in a discrete-time hazard model. In fact, equation (12) above is the log-likelihood for the hazard model in the special case where $h_{i(k+1)}=y_{ik}$ for all k—that is, where the items are strictly ordered and each item is conditional on the prior item.[3] In addition, in the case where $h_{ik} = 1$ by definition for all k (where there are no gate items; each individual responds to all questions), then (12) above is the likelihood for the fixed effects Rasch model with K items per person.

We note that γ_k is interpretable as the "conditional severity" of item k, and θ_i is the value of the latent trait of individual i. The conditional severity of item k corresponds to the value of θ_i at which a person would have a 0.5 probability of answering item k affirmatively, conditional on that person having satisfied the gate conditions for item k. Expressed differently, the conditional severity of item k is the log-odds of a negative response to item k for a person with a value of $\theta_i = 0$ (an "average" person), conditional on that person having met the gate conditions for item k.

The fixed effects model can be estimated using standard software for logistic regression under the model

$$\eta_{ik} = \theta_i + \sum_{j=1}^{K} \gamma_j D_{ij}, \tag{13}$$

where D_{ij} is a dummy variable indicating that person i is responding to item j. One limitation of the fixed effects Rasch model is that we must discard all information from individuals who either endorse all items or who fail to endorse any items. This means that we obtain estimates of θ_i for only the subset of individuals who endorse some, but not all, items.

---

[3]The discrete-time hazard model and the continuation ratio model for ordinal data are formally equivalent models, and both can be seen as a special case of our model, corresponding to a gate matrix where all entries below the diagonal are equal to 1, and all entries on or above the diagonal are 0. In the usual discrete-time and continuation ratio models, we drop θ_i from the model in equation (12), since there is no information from which to estimate it, though we keep θ_i in the case of the multilevel versions of these models (Barber et al. 2000; Reardon, Brennan, and Buka 2002).

4.4. *Random Effects Likelihood*

Under the random effects model, we assume that

$$\theta_i \sim N(0, \tau), \tau \geq 0. \tag{14}$$

The parameters to be estimated are then $(\gamma_1, \ldots, \gamma_K, \tau)$ and the random effects likelihood[4] for participant i is

$$L(\mathbf{Y}_i; \boldsymbol{\gamma}, \tau) = (2\pi\tau)^{-1/2} \int e^{l_{fi} - \theta_i^2/2\tau} d\theta_i. \tag{15}$$

The integral is not available in closed form but can be approximated in several ways, including adaptive Gauss-Hermite Quadrature (Pinheiro and Bates 1995) and the Laplace method (Raudenbush, Yang, and Yosef 2000).

Under the random effects model, person-specific inferences are readily obtained from the posterior distribution of the latent variable θ_i given the data $\mathbf{Y}_i$ and the parameter estimates $(\hat{\gamma}, \hat{\tau})$. The point estimate and uncertainty estimate for the person-specific latent trait are obtained, respectively, from the posterior mean

$$\begin{aligned} \theta_i^* &= E(\theta_i | \mathbf{Y_i}, \hat{\gamma}, \hat{\tau}) \\ &= \int \theta_i \, p(\theta_i | \mathbf{Y_i}, \hat{\gamma}, \hat{\tau}) d\theta_i \\ &= [L(\mathbf{Y}_i; \hat{\gamma}, \hat{\tau})]^{-1} (2\pi\hat{\tau})^{-1/2} \int \theta_i e^{\hat{l}_{fi} - \theta_i^2/2\hat{\tau}} d\theta_i, \end{aligned} \tag{16}$$

and the posterior variance

$$\begin{aligned} V_i^* &= E\left[\left(\theta_i - \theta_i^*\right)^2 | \mathbf{Y_i}, \hat{\gamma}, \hat{\tau}\right] \\ &= \int \left(\theta_i - \theta_i^*\right)^2 p(\theta_i | \mathbf{Y_i}, \hat{\gamma}, \hat{\tau}) d\theta_i \\ &= [L(\mathbf{Y_i}; \hat{\gamma}, \hat{\tau})]^{-1} (2\pi\hat{\tau})^{-1/2} \int \left(\theta_i - \theta_i^*\right)^2 e^{\hat{l}_{fi} - \theta_i^2/2\hat{\tau}} d\theta_i, \end{aligned} \tag{17}$$

where $\hat{l}_{fi} = l_{fi}$ evaluated at $\gamma = \hat{\gamma}$. These posterior means and variances are obtained from integrals approximated via the Laplace method

---

[4]Note that in the special case where $h_{ik}=1$ by definition for all items (where none of the items are conditional on other items), the likelihood in (15) is identical to that of the random effects Rasch model with K items (Raudenbush, Johnson, and Sampson 2003).

(Raudenbush et al. 2005). They are regarded as empirical Bayes esti-
mates because they are conditional upon point estimates of γ and τ.
Unlike the fixed effects model, the random effects model allows us to
retain all persons in the analysis, including those who answered all ques-
tions negatively or all questions affirmatively. Another advantage is that
the random effects model summarizes heterogeneity between persons in
a single parameter (τ), allowing efficient estimation even in the presence
of item missing data under the comparatively mild assumption that the
data are missing at random (Little & Rubin 2002). Finally, the ran-
dom effects model extends naturally to the case where person traits are
modeled as a function of covariates, a procedure we illustrate below.

5. AN EMPIRICAL EXAMPLE

To fit the partially conditional item-response model, we first construct
a person-item data set based on the observed response patterns. We
illustrate the construction of the person-item data set using hypothet-
ical data from three observed subjects. Table 2(a) describes observed
responses to the four alcohol use questions for three hypothetical sub-
jects. Subject 1, a 17-year-old, reported that he had not had a drink in
the last year, and so did not answer the remaining three questions. Sub-
ject 2, a 16-year-old, reported 3 to 5 drinking occasions in the last year.
Because she reported at least one drinking occasion in the last year, she
was also asked how many times she had been drunk in the last year, and
how many times she had had a drink in the last month. She responded
negatively to each of these, meaning that she was not asked the fourth
question. Finally, subject 3, an 18-year-old, reported drinking 25 to 50
times and being drunk 5 times in the last year, and drinking 3 to 5 times
and binge drinking 3 times in the last month.

    The information in these four ordinal items is recoded into 11
binary items, as described above (see Table 1). Table 2(b) reports the
observed item response pattern for these 11 items, with skipped items
coded as (structural) zeros.

    Converting the binary item response pattern in Table 2(b) to a
person-item data set is analogous to constructing a person-period data
set for use with a discrete-time hazard model (Singer and Willett 2003),
although it requires reference to the gate matrix G in order to construct
it. The person-item data set has 11 item indicator dummy variables, one
for each of the 11 binary items in our data. In addition, the person-item
data set includes a binary response variable y, which indicates whether a

TABLE 2(a)
Observed Responses for Three Hypothetical Subjects

ID	# Times Had a Drink, Last Year	# Times Drunk, Last Year	# Times Had a Drink, Last Month	# Times Had 5+ Drinks, Last Month	Sex	Age
1	0 times	-skip-	-skip-	-skip-	M	17
2	3–5 times	0 times	0 times	-skip-	F	16
3	25–50 times	5 times	3–5 times	3 times	M	18

TABLE 2(b)
Binary Item Response Pattern for Three Hypothetical Subjects

ID	Item Number											Sex	Age
	1	2	3	4	5	6	7	8	9	10	11	Sex	Age
1	0	0	0	0	0	0	0	0	0	0	0	M	17
2	1	0	0	0	0	0	0	0	0	0	0	F	16
3	1	1	1	1	0	0	1	1	0	1	1	M	18

TABLE 2(c)
Person-Item Data Set for Three Hypothetical Subjects

ID	y	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9	D_{10}	D_{11}	Sex	Age
1	0	1	0	0	0	0	0	0	0	0	0	0	M	17
2	1	1	0	0	0	0	0	0	0	0	0	0	F	16
2	0	0	1	0	0	0	0	0	0	0	0	0	F	16
2	0	0	0	0	1	0	0	0	0	0	0	0	F	16
2	0	0	0	0	0	0	0	1	0	0	0	0	F	16
3	1	1	0	0	0	0	0	0	0	0	0	0	M	18
3	1	0	1	0	0	0	0	0	0	0	0	0	M	18
3	1	0	0	1	0	0	0	0	0	0	0	0	M	18
3	1	0	0	0	1	0	0	0	0	0	0	0	M	18
3	0	0	0	0	0	1	0	0	0	0	0	0	M	18
3	1	0	0	0	0	0	0	1	0	0	0	0	M	18
3	1	0	0	0	0	0	0	0	1	0	0	0	M	18
3	0	0	0	0	0	0	0	0	0	1	0	0	M	18
3	1	0	0	0	0	0	0	0	0	0	1	0	M	18
3	1	0	0	0	0	0	0	0	0	0	0	1	M	18

subject responded affirmatively to the item represented by each specific line of data.

In the person-item data, there is one observation for each item k for each person i where $h_{ik}=1$. Thus, in Table 2(c), there is a single line of data for subject 1, corresponding to item 1 (*had at least 1 drink in the*
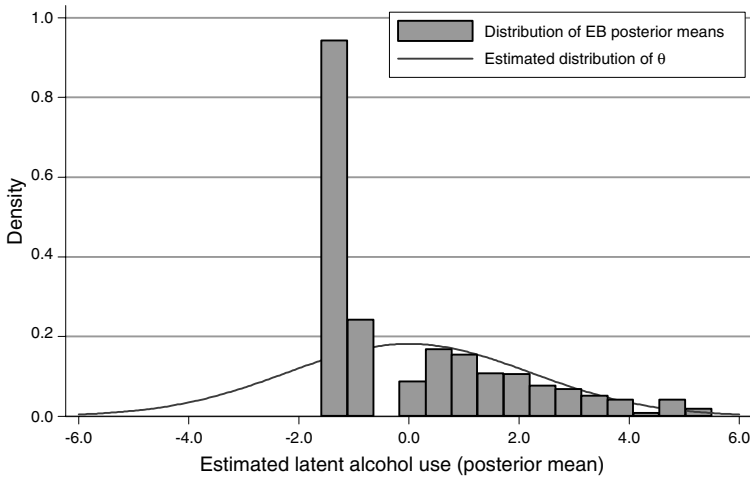
FIGURE 1. Distribution of estimated latent alcohol use (empirical Bayes estimates).

last year), since he was in the risk set for item 1 (everyone is in the risk set for item 1, because it is not conditional on any other item—it has no gate items), but not in the risk set for any other item. Likewise, there are four lines of data for subject 2, one for item 1, to which she answered affirmatively, and one each for items 2, 4, and 7. Because she responded negatively to each of these, she is in the risk set for none of the remaining items, and so none are included in the person-item data. Finally, subject 3 has 10 lines of data in the person-item data, since he was in the risk set for all but one of the questions—he responded negatively to item 5 (*was drunk at least 6 times in the last year*), and so was not in the risk set for item 6 (*was drunk at least 25 times in the last year*).

We fit a random effects model (equations 13–15) to these data using the EM algorithm with Laplace approximation to the likelihood (Raudenbush, Yang, and Yosef 2000) using the software package HLM6 (Raudenbush et al. 2005). From the fitted model we obtain four quantities of interest: (1) person-specific, interval-scale measures of a latent trait (or traits), along with person-specific standard errors of measurement; (2) estimates of conditional and marginal severities for each item in the protocol; (3) person-specific conditional and marginal probabilities of an affirmative response to each item in the protocol; and (4) measures of item information and total survey information. We describe how to obtain and interpret these quantities in the following sections.

5.1. *Person-Specific Estimates of Latent Alcohol Use*

From the fitted random effects model, we obtain empirical Bayes person-specific posterior means (θ^*_i) and variances (V^*_i) as given by equations (16) and (17). Figure 1 illustrates the distribution of θ^*_i, $i = 1, \ldots, n$. The fitted random effects model also yields an estimate of the variance τ of the latent alcohol use trait ($\hat{\tau} = 4.816$), and the solid line in Figure 1 describes this estimated distribution of the latent trait. The distribution of θ^*_i has a large spike at $\theta^*_i = -1.58$. In fact, 647 of the 1452 subjects have $\theta^*_i = -1.58$; these are those individuals who responded *no* to the first item (*ever drink in the last year*). Because these individuals were asked no other items, their estimated θ^*_i is based on a single item, and so contains much more uncertainty than does θ^*_i for individuals who were asked more questions.

Figure 2 illustrates the precision of θ^*_i, given by the posterior standard deviation $(V^*_i)^{1/2}$. The left-vertical describes the posterior standard deviation in the units of θ; the right-vertical axis describes the posterior standard deviation expressed in terms of the estimated standard deviation of θ, $(V^*_i/\hat{\tau})^{1/2}$; this conversion makes the magnitude of the posterior standard deviations more interpretable. The 647 subjects who reported never drinking in the last year have the largest posterior standard
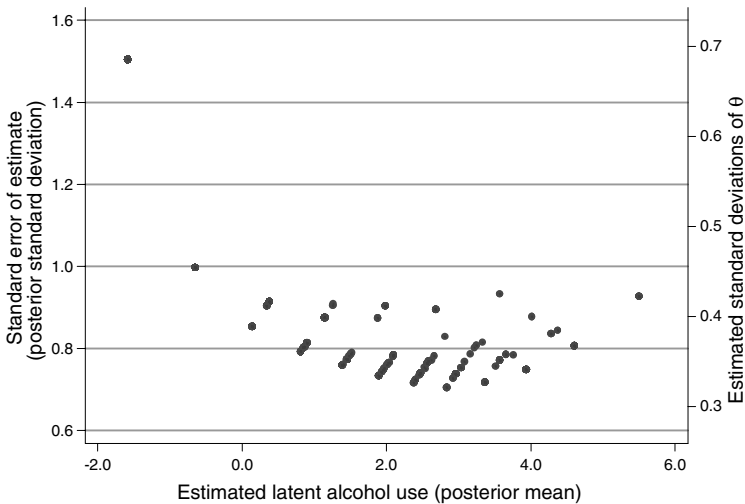

FIGURE 2. Precision of estimated latent alcohol use.

TABLE 3
Estimated Conditional and Marginal Item Severities

Item	Description	Estimated Conditional Severity		Estimated Marginal Severity
1	Alcohol in year	−0.362	(0.089)	−0.362
2	Alcohol 6x in year	1.388	(0.116)	1.656
3	Alcohol 25x in year	2.823	(0.150)	3.215
4	Drunk in year	0.481	(0.114)	1.003
5	Drunk 6x in year	3.423	(0.152)	3.557
6	Drunk 25x in year	4.195	(0.223)	4.777
7	Alcohol in month	0.729	(0.113)	1.169
8	Alcohol 3x in month	2.133	(0.144)	2.577
9	Alcohol 10x in month	4.666	(0.202)	4.846
10	Binge in month	2.081	(0.138)	2.543
11	Binge 3x in month	3.607	(0.186)	4.017

Note: Standard errors in parentheses. Conditional difficulties are estimated coefficients from fitted model; marginal difficulties are the value of θ that corresponds to a marginal probability of 0.5, given the estimated conditional item difficulties and the gate matrix.

deviations $((V_i^*)^{1/2} = 1.50 = .69\hat{\tau}^{1/2})$, reflecting the lack of information from which to estimate θ_i for these individuals. The precision is greatest for individuals with values of θ_i^* in the middle range of the estimates—corresponding to individuals whose latent use is well-differentiated by the items on the survey. We return to this point in the discussion of the survey information function below.

5.2. *Estimated Item Conditional and Marginal Severity*

The fitted random effects model yields the estimated parameters shown in Table 3. The estimated coefficients on the 11 item indicator variables represent the estimated item conditional severity for each of the items. As we noted above, the conditional severity of item k is interpreted as the value of θ_i at which an individual would have a 0.5 probability of answering item k affirmatively, given that he or she is in the risk set for the item.

In addition to obtaining estimates of the item conditional severities, we can also obtain estimates of the item marginal severities. The
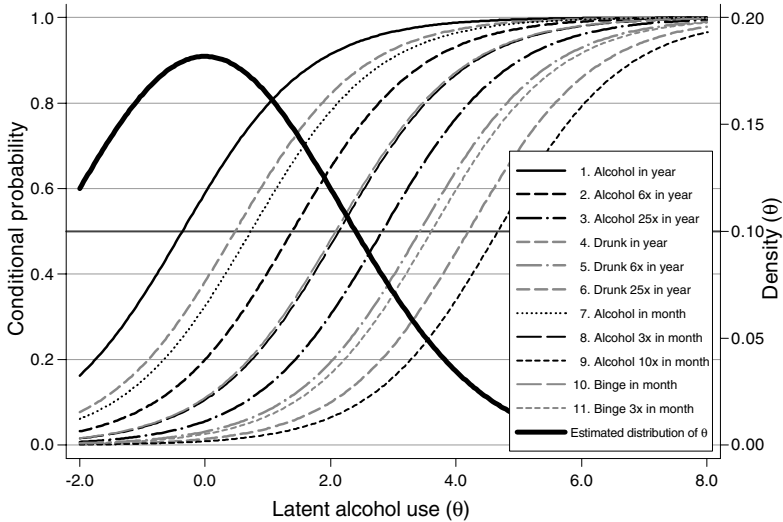
FIGURE 3. Estimated item conditional probabilities, by latent trait.

marginal severity of item k is the value of θ_i at which an individual would have a 0.5 probability of answering item k affirmatively, given the estimated item conditional severities and the gate matrix. These marginal severities are of more interest—and are more interpretable—than the conditional severities, since they correspond to estimated behavioral prevalences in the population. Given the estimated item conditional severity parameters from Table 3, we compute the item marginal severities by finding the value of θ such that $\pi_k = 0.5$, where π_k is given by

$$\pi_k = \left(1 + e^{-(\theta - \hat{\gamma}_k)}\right)^{-1} \prod_{j=1}^{K} \left(1 + e^{-(\theta - \hat{\gamma}_j)}\right)^{-G_{kj}} \tag{18}$$

Since π_k is a strictly increasing continuous function of θ, we solve $\pi_k(\theta) = 0.5$ for θ by interpolation. Table 3 reports these estimated item marginal severities.

Figures 3 and 4 illustrate the fitted conditional and marginal probability curves for each item. Figure 3 illustrates the conditional probability curve for each item, with the estimated distribution of
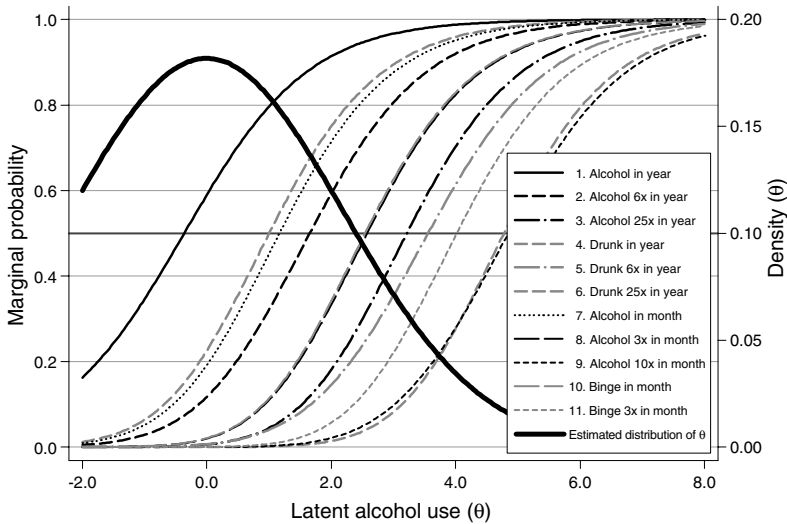
FIGURE 4. Estimated item marginal probabilities, by latent trait.

θ_i overlaid. Note that the item conditional severity parameters reported in Table 3 and illustrated in Figure 3 should not be interpreted as providing a meaningfully ordered ranking of item behavioral severity. For example, the conditional severity of item 11 (*had 5 or more drinks in a row at least 3 times in the last month*) is slightly higher than the conditional severity of item 5 (*was drunk at least 6 times in the last year*), but we cannot tell from these coefficient estimates alone whether the prevalence of binge drinking three or more times in the last month is greater or less than the prevalence of being drunk at least six times in the last year. This is because the conditional severities depend on the item structure (the gate matrix) and so are not straightforwardly interpretable. We note also that, in our example, the estimated conditional severities are all greater than 0, except that for item 1 (*ever drank in the last year*). This implies that we have little information in the data with which to discriminate individuals with low values of θ_i from one another, a point we return to below.

Figure 4 illustrates the estimated marginal probability curves, computed from equation (18). Unlike the item conditional probability curves, the marginal probability curves are not parallel, since they are the product of multiple conditional probability curves. However, the marginal probability curves do yield a meaningful ordered ranking of

item severity. As we expect given the item structure, the least severe item is item 1 (*ever have a drink in the last year*); next are items 3 (*ever drunk in the last year*) and 4 (*ever have a drink in the last month*). At the other end of the ranking, the highest severity are items 9 (*had a drink 10 or more times in the last month*) and 6 (*was drunk at least 25 times in the last year*). Note that the marginal probability curves for these two items cross. Although the curves are so close in this case as to be statistically indistinguishable (particularly given the small number of individuals in the risk set for either item), the curve crossing—were it significant—would be interpreted as indicating that item 6 is the most severe item among those with values of θ_i above 4.0, while item 9 is the most severe among those with lower values of θ_i.

5.3. *Estimated Person-Specific Item Conditional and Marginal Probabilities*

Given the estimated item conditional severity parameters from Table 3 and the person-specific (empirical Bayes) estimates of θ_i, we can compute estimated person-specific conditional probabilities for each item. These are given by substituting the estimated conditional severity parameters and the estimated $\hat{\theta}_i^*$ from equation (16) into equation (10):

$$\hat{\phi}_{ik} = \left(1 + e^{-\left(\theta_i^* - \hat{\gamma}_k\right)}\right)^{-1}. \tag{19}$$

From these estimates, we can compute the predicted person-specific marginal probability for each item k as

$$\hat{\pi}_{ik} = \hat{\phi}_{ik} \prod_{j=1}^{K} \hat{\phi}_{ij}^{G_{kj}}. \tag{20}$$

5.4. *Comparing Observed and Predicted Item Conditional and Marginal Probabilities*

In addition to examining the fitted conditional and marginal probability curves, we wish to examine the fit of the model to the observed data.

One way of doing this is to assess how closely the model predictions fit the observed conditional and marginal proportions of individuals endorsing each item. We have two methods of checking this. The first relies on predicting the person-specific conditional and marginal probabilities as above, and then averaging these over those in the risk set for a given item. Specifically, we first compute each $\hat{\phi}_{ik}$ and $\hat{\pi}_{ik}$ from equations (19) and (20), and then, for each item k, we compute $\bar{\hat{\phi}}_k = \sum_i h_{ik} \hat{\phi}_{ik}$, the predicted average conditional probability of endorsing item k among those in the risk set for item k, and $\bar{\hat{\pi}}_k = \sum_i \hat{\pi}_{ik}$, the predicted average marginal probability of endorsing item k among those in the sample. We then compare these predictions with the observed conditional and marginal proportions in the data (see Table 4).

Because of shrinkage in θ_i^* as an estimate of θ_i, the predicted average conditional probabilities will be biased, though the direction of bias is not consistent. Persons with high true values of θ_i will be assigned estimates θ_i^* that are biased negatively (toward zero), thus underestimating their conditional probabilities ϕ_{ik}. In contrast, persons with low true values of θ_i will be assigned estimates θ_i^* that are biased positively (toward zero), thus overestimating their conditional probabilities ϕ_{ik}. For any given item k that is conditional on at least one other item, the risk set R_k will, in general, contain more individuals with high values of θ_i than with low values of θ_i, since those with higher values of θ_i will be more likely to respond affirmatively to the gate items for k. Thus, when we average the $\hat{\phi}_{ik}$ over those in R_k, the resulting average will tend to be a negatively biased estimate of the true average conditional probability of item k. Even if these biases are small, when these biased conditionals are multiplied to obtain marginal probabilities, the negative biases will tend to accumulate, yielding larger bias in the estimated average marginal probabilities than in the estimated marginal conditional probabilities. This pattern is evident in Table 4.

The second method of assessing the model fit relies on integrating the estimated marginal probability curve over the estimated distribution of θ in order to obtain estimated average conditional and marginal probabilities. Specifically, we compute the estimated average marginal probability for each item k as

$$\bar{\hat{\pi}}_k = (2\pi\hat{\tau})^{-1/2} \int \hat{\pi}_{k\theta} e^{-\theta^2/2\hat{\tau}} d\theta, \qquad (21)$$

TABLE 4
Observed and Predicted Conditional and Marginal Probabilities, Method 1

Item	Conditional Probability			Marginal Probability		
	Observed	Predicted	Difference	Observed	Predicted	Difference
1. Alcohol in year	.554	.518	−.036	.554	.518	−.036
2. Alcohol 6x in year	.472	.461	−.011	.262	.227	−.035
3. Alcohol 25x in year	.445	.438	−.007	.116	.097	−.019
4. Drunk in year	.615	.616	.001	.341	.296	−.045
5. Drunk 6x in year	.277	.263	−.014	.094	.083	−.011
6. Drunk 25x in year	.401	.393	−.008	.038	.030	−.008
7. Alcohol in month	.576	.573	−.003	.320	.277	−.043
8. Alcohol 3x in month	.494	.490	−.004	.158	.138	−.020
9. Alcohol 10x in month	.249	.233	−.016	.039	.031	−.008
10. Binge in month	.502	.499	−.003	.160	.141	−.019
11. Binge 3x in month	.399	.392	−.007	.064	.056	−.008

where $\hat{\pi}_{k\theta}$ is the estimated marginal probability at θ, as defined in equation (20).[5] From (21), we compute the estimated average conditional probability of endorsing item k among those in the risk set for item k as

$$\bar{\bar{\phi}}_k = \frac{\bar{\bar{\pi}}_k}{\bar{\bar{\pi}}_{R_k}}, \tag{22}$$

where $\bar{\bar{\pi}}_{R_k}$ is the estimated average marginal probability of being in the risk set for item k. In the case where each item has at most only a single immediately prior gate item (that is, for each item k, there is at most one item j such that $y_{ij} = 1$ implies $h_{ik} = 1$), the marginal probability of being in the risk set for item k is simply the marginal probability of endorsing the single immediately prior gate item.[6] Table 5 compares the predicted average conditional and marginal probabilities with the observed conditional and marginal proportions in the data.

Note that method one (Table 4) yields very good estimates of the average conditional probabilities, while method two (Table 5) yields very good estimates of the marginal probabilities. This difference is likely due to the fact that method one is based on directly estimating the average conditional probabilities from the model estimates and then multiplying these to obtain the average marginals. Here slight biases in the estimated conditional probabilities (particularly if the biases are generally in the same direction, as they are here) will be multiplied into larger biases in the estimated marginal probabilities in this case. Method two, in contrast, is based on directly estimating the average marginal probabilities and then taking ratios of these to compute the average conditionals. Again, slight biases in the estimated marginal probabilities may be

---

[5]The integral in (21) has no closed form, so we can evaluate it numerically, or by simulating a distribution of $\theta \sim N(0,\tau)$ and averaging the value of π_{ik} over the simulated distribution of θ.

   [6]In the more general case, as for example, in the gate matrix described by matrix (5), the marginal probability of being in the risk set for item k is given by

$$\pi_{R_k} = \prod_{m_1=1}^{K} \prod_{m_2=1}^{K} \cdots \prod_{m_K=1}^{K} (\pi_{m_1})^{G[k,m_1]} \Big[(\pi_{m_2})^{G[m_1,m_2]} \cdots \Big[(\pi_{m_K})^{G[m_{K-1},m_K]}\Big]^{-1} \cdots \Big]^{-1}.$$

TABLE 5
Observed and Predicted Conditional and Marginal Probabilities, Method 2

Item	Conditional Probability			Marginal Probability		
	Observed	Predicted	Difference	Observed	Predicted	Difference
1. Alcohol in year	.554	.551	−.003	.554	.551	−.003
2. Alcohol 6x in year	.472	.469	−.003	.262	.259	−.003
3. Alcohol 25x in year	.445	.400	−.045	.116	.103	−.013
4. Drunk in year	.615	.618	.003	.341	.341	.000
5. Drunk 6x in year	.277	.260	−.017	.094	.089	−.005
6. Drunk 25x in year	.401	.338	−.063	.038	.030	−.008
7. Alcohol in month	.576	.578	.002	.320	.318	−.002
8. Alcohol 3x in month	.494	.480	−.014	.158	.153	−.005
9. Alcohol 10x in month	.249	.208	−.041	.039	.032	−.007
10. Binge in month	.502	.489	−.013	.160	.156	−.004
11. Binge 3x in month	.399	.363	−.036	.064	.057	−.007

compounded into larger discrepancies in the conditional probabilities through multiplication. Since the marginal probabilities are generally likely to be of greater interest, method two provides a more useful assessment of model fit.

In both cases, the discrepancies tend to be negative, meaning that the model appears to slightly underestimate the observed conditional and marginal probabilities of item endorsement. The greatest discrepancies between the observed and predicted probabilities occur for the rarest items, possibly because the upper tail of the true distribution of θ is somewhat longer than that given by the simulated normal distribution. This is also suggested by the apparent correlation between the bias in the predicted conditional probabilities and the marginal probabilities of the items—if the predicted distribution is too thin in the upper tail, this will show up most dramatically in computing the conditional probabilities for the items with the highest conditional severities. Overall, however, the model appears to fit the data rather well.

5.5. *Computing Item and Survey Information*

As we note above, Figures 2, 3, and 4 each provide some insight into the information content of the survey. It is evident from Figure 2 that the uncertainty in the estimated latent alcohol use trait is smallest for θ_i^* in the vicinity of 3.0. Likewise, Figures 3 and 4 suggest that the survey items will discriminate best among individuals with values of θ somewhere in the range of 1.0 to 4.0, since most of the items have conditional and marginal severities in this range. The items are not independent of one another, so—unlike in a Rasch model, where all items are independent of one another—the total information in the survey does not necessarily correspond to the density of the item severities, since more severe items will be asked of relatively few subjects.

We formalize the notion of survey information here. We consider two approaches to defining the survey information. First, we consider Fisher information based on the fixed effects model. The fixed effects model yields a simple, interpretable, and useful measure of total survey information. Second, we compute a person-specific measure of information: the inverse of the empirical Bayes posterior variance V_i^* of θ_i given the data and the estimates of γ and τ. A graph of this measure of information as a function of the person-specific empirical Bayes

estimates θ_i^* provides additional insight about the values of θ for which the survey is most informative.

Under the fixed effects model, the observed Fisher information for person i is defined as the negative of the second derivative of the log-likelihood, given θ_i:

$$
\begin{aligned}
I_i(\theta_i) &= -\frac{\partial^2 l_{fi}}{\partial \theta^2} \\
&= -\frac{\partial^2}{\partial \theta^2} \sum_{k=1}^{K} h_{ik} \left[y_{ik} \ln \left(\frac{\phi_{ik}}{1 - \phi_{ik}} \right) + \ln(1 - \phi_{ik}) \right] \quad (23) \\
&= \sum_{k=1}^{K} h_{ik} \phi_{ik} (1 - \phi_{ik}).
\end{aligned}
$$

From this, we can derive the expected Fisher information, given θ_i, as

$$
\begin{aligned}
E[I_i(\theta_i)] &= E \left(\sum_{k=1}^{K} h_{ik} \phi_{ik} (1 - \phi_{ik}) \right) \\
&= \sum_{k=1}^{K} E(h_{ik}) \phi_{ik} (1 - \phi_{ik}) \\
&= \sum_{k=1}^{K} \pi_{i R_k} \phi_{ik} (1 - \phi_{ik}) \\
&= \sum_{k=1}^{K} \pi_{ik} (1 - \phi_{ik}).
\end{aligned} \quad (24)
$$

As equation (24) indicates, the expected information content of the survey protocol, given θ_i, is a sum of the expected information from each item. The information contained in each item is the product of the expected size of the risk set for the item and the expected information content of the item.[7] Items will provide the most information for values of θ such that the risk set, given θ, is large, and the conditional probability of the item, given θ, is 0.5. This is useful for considering how to construct survey protocols designed to measure latent traits.

The Fisher information is based on the fixed effects model—that is, it describes the information conditional on θ. The Bayesian information is given by the inverse of the posterior variance in equation (17).

---

[7]We note that, in the case of the Rasch model, all individuals are in the risk set for each item, so the expected Fisher information given in (24) reduces in this special case to $\sum_{k=1}^{K} \phi_{ik}(1 - \phi_{ik})$, as expected.
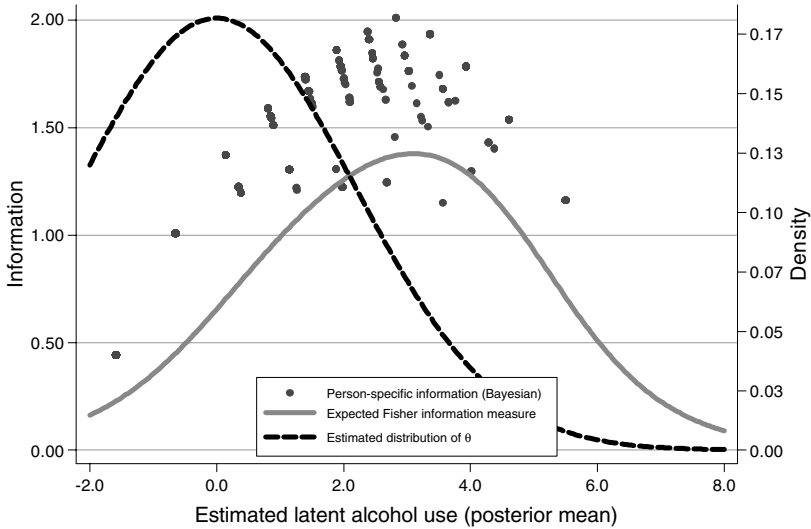
FIGURE 5. Estimated total survey information, by latent trait.

Figure 5 plots both the estimated Bayesian information ($\hat{V}_i^{*-1}$) and the expected Fisher information (equation 24) against θ_i. The figure shows that the total information based on the fixed effects model (solid curve) is highest just between 3.0 and 4.0. The graph of the empirical Bayes person-specific information (see the dots in the figure) tells a very similar story. The maximum values of this information measure are also between 3.0 and 4.0. As we might expect, the empirical Bayes information is uniformly higher than the information from the fixed effects model. The information at any θ from the fixed effects model is based only on the data for a single person, while the empirical Bayes information is the sum of that person-specific information and the information about γ and τ provided by the rest of the sample. In this way, the empirical Bayes procedure augments the information from each participant, "borrowing strength" from the sample as a whole.

It is evident from Figure 5 that the survey has little discriminatory power for individuals with $\theta_i < 0$. It may be that we are not interested in differentiating among such individuals, but if we were, Figure 5 would suggest the need to include more low severity items in the survey protocol.

6. ASSESSING DIMENSIONALITY

Dimensionality of the latent characteristic may be assessed in at least
three ways. First, we may hypothesize multiple dimensions of latent
alcohol use and attempt to confirm that different items correspond
to different latent dimensions by examining the estimated correlation
matrix among the hypothesized dimensions. Second, we may investi-
gate whether observed covariates are similarly associated with the hy-
pothesized dimensions, reasoning that if covariates are differently as-
sociated with the hypothesized dimensions, then they cannot be con-
sidered unidimensional (Raudenbush, Johnson, and Sampson 2003).
Third, we may employ a differential item functioning test to investigate
whether all the items are similarly associated with observed covariates.
We describe each of these approaches to investigating dimensionality
below.

6.1. *Examining the Correlation Matrix Among Hypothesized Dimensions*

Let θ^a and θ^b denote latent characteristics, and let a_k and b_k be dummy
variables indicating whether item k is an indicator of latent characteristic
θ^a or θ^b, respectively (we assume $a_k + b_k = 1$ for all k). Then we fit the
random effects model

$$\ln\left(\frac{\phi_{ik}}{1 - \phi_{ik}}\right) = a_k\theta_i^a + b_k\theta_i^b + \sum_{j=1}^{K} \gamma_j D_{ij},$$

$$\begin{bmatrix} \theta_i^a \\ \theta_i^b \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_a & \tau_{ab} \\ \tau_{ab} & \tau_b \end{bmatrix}\right), \tag{25}$$

where D_{ij} is a dummy variable indicating that a response refers to item
j for person i. The estimated correlation between θ^a and θ^b is indicative
of the extent to which the items measuring θ^a and θ^b represent different
dimensions.

6.2. *Examining the Association of Covariates with Hypothesized Dimensions*

If θ^a and θ^b measure the same dimension, then an observed covariate X should be similarly associated with both. To test this, we fit the model

$$\ln\left(\frac{\phi_{ik}}{1-\phi_{ik}}\right) = a_k\left(u_i^a + \gamma_a X_i\right) + b_k\left(u_i^b + \gamma_b X_i\right) + \sum_{j=1}^{K}\gamma_j D_{ij},$$

$$\begin{bmatrix} u_i^a \\ u_i^b \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_a & \tau_{ab} \\ \tau_{ab} & \tau_b \end{bmatrix}\right) \qquad (26)$$

and test the hypothesis $H_0: \gamma_a = \gamma_b$. A rejection of H_0 indicates that θ^a and θ^b measure different dimensions.

6.3. *Assessing Dimensionality via Differential Item Functioning*

A third way of assessing dimensionality is to examine the model for differential item functioning (DIF). If the items all measure the same latent characteristic, then observable covariates ought to be similarly associated with each of the items. We can test for DIF by each observed covariate X by fitting the model

$$\ln\left(\frac{\phi_{ik}}{1-\phi_{ik}}\right) = \theta_i + \sum_{j=1}^{K}\left(\gamma_j + \delta_j X_i\right) D_{ij}, \quad \theta_i \sim N(0, \tau) \qquad (27)$$

and testing the null hypothesis $H_0: \delta_1 = \delta_2 = \ldots = \delta_K$ (We use a likelihood ratio test, comparing this model's deviance to the deviance of a model where the δ_js are constrained to be equal.)

6.4. *Empirical Assessment of Substance Use Dimensionality*

To illustrate these different approaches to assessing dimensionality, we use a slightly different example than above. Here we use the same sample of adolescents from the PHDCN sample but include information from two additional survey questions—measuring marijuana use—in the data. Specifically, we use information from two marijuana use items:

5. *How many times did you use marijuana in the last year?* (possible responses were: never, 1–2 times, 3–5 times, 6–11 times, 12–24 times, 25–50 times, 51–99 times, 100–199 times, 200 or more times).

6. *How many times did you use marijuana in the last month?* (asked of those whose response to question 5 indicated they used marijuana at least once in the last year; possible responses were: never, 1–2 times, 3–5 times, 6–9 times, 10–14 times, 15–20 times, 21 or more times).

As with the four alcohol items, from these two items we construct multiple binary items, each indicating whether an individual's response is above a certain threshold (we use thresholds of 1, 6, 25, and 100 for the first item, and thresholds of 1, 3, and 10 for the second). The full set of 18 derived alcohol and marijuana binary items and their response patterns are described in Table 6, and their item structure and corresponding gate matrix is given below (where the items are ordered as above):
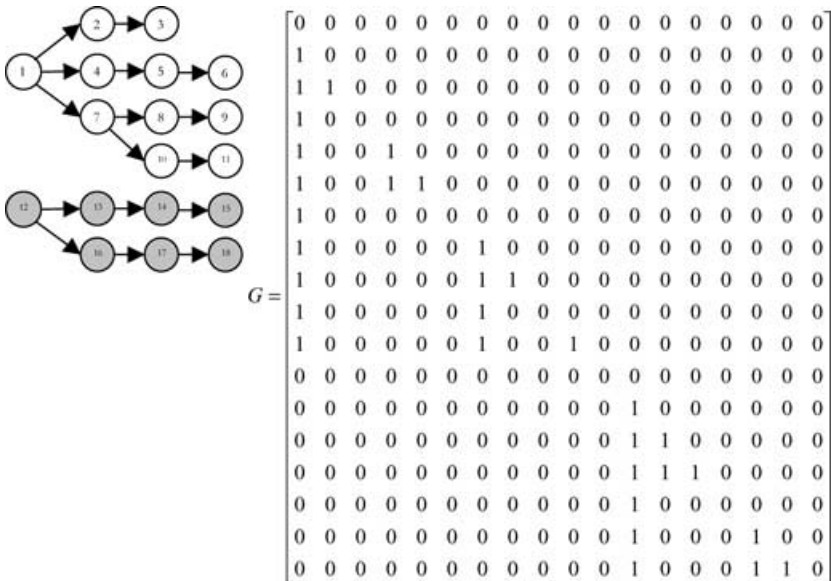

$$
G =
\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0
\end{bmatrix}
$$

(28)

TABLE 6
Observed Alcohol and Marijuana Item Response Frequencies

Item	Description	Gate Items	Marginal Probability	Risk Set	Number of *Yes*	Conditional Probability
1	Alcohol in year	-none-	.557	1,431	797	.554
2	Alcohol 6x in year	1	.261	797	374	.472
3	Alcohol 25x in year	1,2	.116	374	166	.445
4	Drunk in year	1	.342	797	489	.615
5	Drunk 6x in year	1,4	.093	489	133	.277
6	Drunk 25x in year	1,4,5	.037	133	53	.401
7	Alcohol in month	1	.319	797	457	.576
8	Alcohol 3x in month	1,7	.156	457	223	.494
9	Alcohol 10x in month	1,7,8	.040	223	57	.249
10	Binge in month	1,7	.160	457	229	.502
11	Binge 3x in month	1,7,10	.064	229	91	.399
12	Marijuana in year	-none-	.293	1,431	419	.293
13	Marijuana 6x in year	12	.164	419	235	.561
14	Marijuana 25x in year	12,13	.094	235	134	.570
15	Marijuana 100x in year	12,13,14	.046	134	66	.493
16	Marijuana in month	12	.177	419	252	.601
17	Marijuana 3x in month	12,16	.112	252	159	.631
18	Marijuana 10x in month	12,16,17	.065	159	93	.585

Note: The sample is slightly smaller here than in Table 1, due to the omission of 21 individuals who did not answer the marijuana items.

We hypothesize that alcohol and marijuana use may be separate dimensions of substance use. Under this assumption, we view the 11 alcohol items (unshaded above) as conditional indicators of a latent alcohol use dimension, and the 7 marijuana items (shaded above) as conditional indicators of a latent marijuana dimension. We might hypothesize that these two latent characteristics are distinct in the population—an individual's alcohol use may not correspond well to his or her marijuana use.

When we fit the model (25) to our data, using each of the hypothesized dimensional structures, we obtain an estimated correlation near 1.0 between the latent alcohol and marijuana use levels. This high correlation is more a result of the fact that we have very little information from which to separately estimate the two factors, rather than strong evidence that they are perfectly correlated. So this is a rather weak test of dimensionality, except in the case where we have far more items than those here.

TABLE 7
Estimated Coefficients on Selected Covariates Predicting Hypothesized Separate
Dimensions of Substance Use

	Male	SES	Age	Black	Hispanic
Main Effect[a]	0.839***	0.160**	0.816***	−0.823***	−1.025***
Dimension-Specific Effect					
Alcohol[b]	0.771***	0.183***	0.985***	−1.208***	−1.098***
Marijuana[b]	0.910***	0.137*	0.649***	−0.466*	−0.988***
p-value[c]	*0.167*	*0.289*	*<.001*	*<.001*	*0.476*
Differential Item					
Functioning (DIF) Test					
p-value[d]	*0.095*	*0.084*	*<.001*	*<.001*	*>.500*

    [a] Asterisks indicate *p*-value from deviance test of null hypothesis that the estimated coefficient is zero (*$p < .05$; **$p < .01$; ***$p < .001$). Coefficient estimates obtained from model that assumes a single substance use domain.
    [b] Asterisks indicate *p*-value from deviance test of null hypothesis that the estimated coefficient on the specified dimension is zero (*$p < .05$; **$p < .01$; ***$p < .001$).
    [c] *p*-value from deviance test of equivalence of coefficients on both dimensions.
    [d] *p*-value from deviance test of equivalence of coefficients on all 18 alcohol and marijuana items.
    Notes: Coefficient estimates are taken from models of the form shown in (18), but with the constraint $u_i^a = u_i^b$ for all i (see text), and with each covariate entered separately in a different model (except *black* and *hispanic*, which are entered in a model together).

    Our second approach to assessing dimensionality derives from the insight that if θ^a and θ^b measure the same dimension of substance use, then an observed covariate X should be similarly associated with both (Raudenbush, Johnson, and Sampson 2003). We test this, using covariates *age*, *sex*, *SES*,[8] and race/ethnicity (*black* and *hispanic*, versus omitted category *white*) in models of the form shown in equation (26).[9] The resulting estimates, and the *p*-values corresponding to the tests of the hypothesis that the associations between X and each of the hypothesized dimensions are equal, are shown in Table 7.
    Table 7 provides some evidence of multidimensionality. First, the relationship between age and alcohol use ($\hat{\gamma}_a = 0.985$, $p < .001$)

    [8]SES is a standardized composite indicator of socioeconomic status, derived from measures of mother and father's education levels and occupations, and family income.
    [9]The fitted model is slightly different than described in Equation (26): this model would not converge when covariates were included on the alcohol and marijuana constructs, because the alcohol and marijuana random effects u^a and u^b were too highly correlated. We constrain the two random effects to be equal to obtain the estimates shown in Table 7.

appears greater than the relationship between age and marijuana use ($\hat{\gamma}_b = 0.649$, $p < .001$). Indeed, the difference between these two coefficients ($0.985\text{-}0.649 = 0.336$) is highly statistically significant ($p < .001$). This means that, in the age range represented in our sample, use levels differ by age more for alcohol than for marijuana. In addition, the difference between blacks and whites in alcohol use ($\hat{\gamma}_a = -1.208$, $p < .001$) appears greater than the difference between blacks and whites in marijuana use ($\hat{\gamma}_b = -0.466$, $p < .05$). In fact, the difference between these two coefficients ($-1.208-(-0.466)= -0.742$) is highly statistically significant ($p<.001$). We cannot, however, reject the null hypothesis that each of the other covariates tested (*sex*, *SES*, and *hispanic*) are similarly associated with the hypothesized alcohol and marijuana dimensions.

Our third approach to assessing dimensionality is to test the items for differential item functioning (DIF). Using the same five covariates as above, we fit models of the form shown in equation (27), and test the null hypothesis that the association of the covariate with each of the 18 alcohol and marijuana items is constant across items. Figure 6 shows the 95% confidence intervals for each coefficient for each item. The solid horizontal line in each part of the figure indicates the estimated coefficient from the model assuming a constant relationship between the covariate and each item (these are the "main effects" reported in Table 7). The two dashed lines in each part represent the estimated coefficients from the model allowing the covariate to have different associations with the latent alcohol and marijuana use traits (these are the "dimension-specific effects" reported in Table 7). Table 7 reports the *p*-values from deviance tests of the null hypothesis of constant association across items. The results of the DIF test are consistent with those of the covariate modeling approach above: both age and race (black versus white) are differentially associated with the conditional log-odds of endorsing different items.

The three approaches to assessing the dimensionality of the item responses described here have different strengths and limitations. The first approach, based on estimating the correlation among hypothesized latent characteristics indicated by different items, allows us to test the hypothesis that distinct subsets of items are indicators of separate dimensions without reference to any specific observable covariate. This test, however, generally has very low power, and so may easily result in a failure to detect existing multidimensionality unless the latent characteristics are estimated with high precision (which requires a large
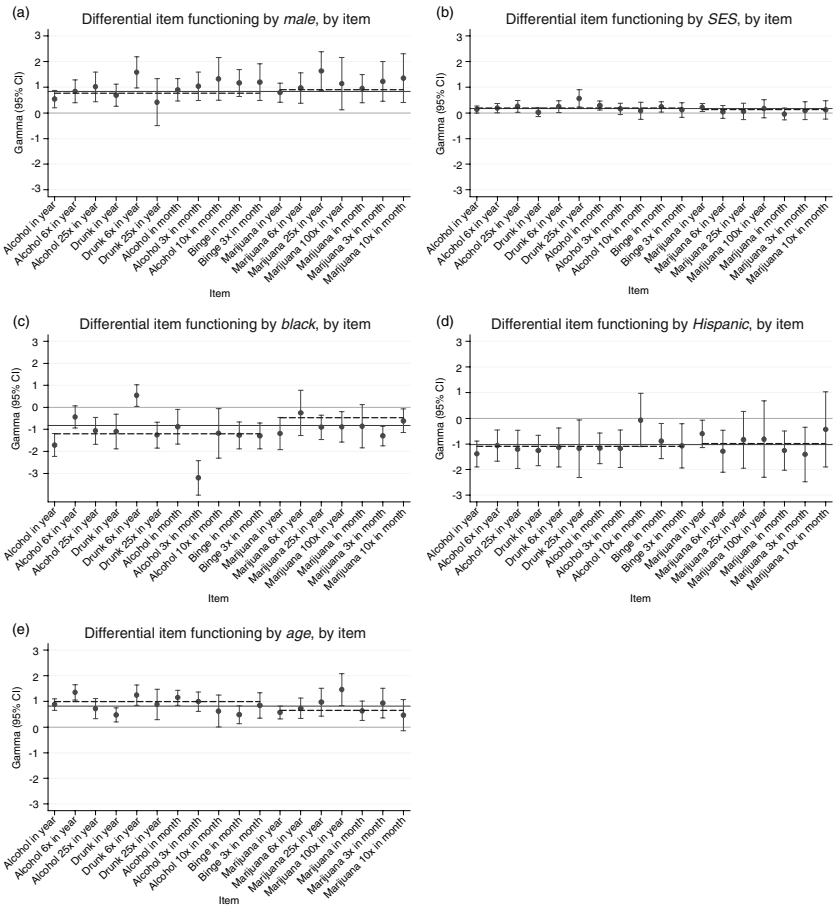
FIGURE 6. Estimated coefficients from differential item functioning tests, by covariate.

number of items to provide information across the range of the latent characteristics' distributions). The second approach, based on testing whether observed covariates are similarly associated with each hypothesized latent characteristic, has greater power to detect hypothesized multidimensionality, but only if that multidimensionality is associated with observable covariates. The third approach, based on the DIF test, has the advantage that it does not require us to specify a hypothesized dimensional structure to the items. The DIF test, however, has low power when the number of items becomes large, since the test will have degrees

of freedom equal to K, the number of items. Moreover, the DIF test, like the second approach, can detect multidimensionality only if that multidimensionality is associated with observable covariates.

Under the assumption of unidimensionality, the latent trait model we propose here, like all logit models, contains an implicit set of proportional odds assumptions. In particular, the model makes assumptions analogous to the level-2 proportional odds and proportional error assumptions described in Reardon, Brennan, and Buka (2002). Applied to our model, these assumptions require that a given difference in a covariate or in the residual unobserved latent trait (that part of the latent trait not explained by the covariates in the model, if any) is associated with a proportional difference in the conditional odds of an affirmative response to each item. One way to view the dimensionality tests described here is as tests of these proportionality assumptions— our approach assumes a definition of dimensionality based on the insight that multidimensionality is evident in a violation of one or more proportional odds assumptions. Expressed differently, the proportional odds assumption is equivalent to the additivity assumption in the Rasch model. One could, in principle, relax this assumption and use a two-parameter model rather than a Rasch model to study the conditional probabilities in order to test the additivity assumption of the Rasch scale and to identify items that do not fit the Rasch scale (Raudenbush, Johnson, and Sampson, 2003).

7. CONCLUSION

We noted above that one approach used by some researchers in modeling nonindependent items like these is to create a variable reflecting the sum or average of the item responses. It is illustrative to compare our estimates of the person-specific latent trait values with those that would be obtained using this simpler sum-of-item-responses approach. Figure 7 illustrates the relationship in our data between the observed and expected sums of alcohol item responses[10] and the estimated latent alcohol use $\hat{\theta}_i^*$ (we use only the 11 alcohol items here, not the additional

---

[10]The expected sum of item responses, given θ_i^*, is simply the sum of the K marginal item probabilities at θ_i^*. We note also that the expected sum of item responses tends to underestimate the observed sum of item responses, particularly for large values of θ_i^*, as a result of shrinkage in the estimation of θ_i^*.
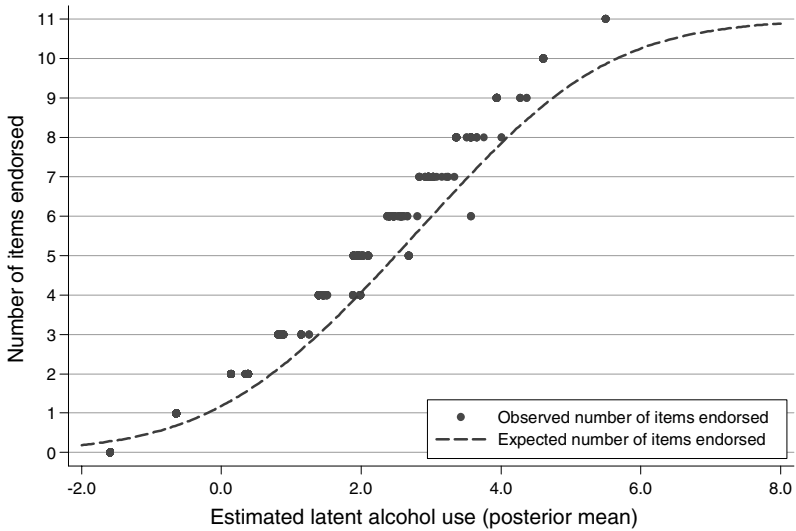
FIGURE 7. Number of items endorsed, by estimated latent alcohol use.

7 marijuana items). The correlation between the observed sum of item responses and $\hat{\theta}_i^*$ is strikingly high (r = .989), as evident in Figure 7. We might reasonably ask, then, what is the value of our modeling approach when it yields estimates that are largely indistinguishable from a simple sum of the item responses?

The latent trait model proposed here has a number of advantages over the sum-of-item-responses approach. First, the latent trait model provides estimates (and associated standard errors) of the item conditional severities, which can be used to estimate item marginal severities. These can be used to compute person-specific marginal probabilities of saying *yes* to a given item, and so provide information regarding the relative prevalence of specific behaviors at any given level of the latent trait. In conjunction with the estimate of τ, these marginal probabilities allow us to estimate the prevalence of specific behaviors in the population. The sum-of-item-responses approach yields no information on item severity or prevalence, which in many cases may be a primary goal of the analysis.

Second, our approach yields not only estimates of the distribution of the latent trait and person-specific estimates of θ_i^*, but also person-specific estimates of the uncertainty in θ_i^* (equation 17). The sum-of-item-responses approach provides no estimate of uncertainty;

it assumes no measurement error. A third advantage of our model is that it is robust to item-missing data, under missing-at-random (MAR) and coarsening-at-random assumptions; see Barber et al. (2000); Little and Rubin (2002). The sum-of-item-responses approach does not allow missing data, since an average of nonmissing items will be biased upward relative to the average of the complete set of items (because the marginal probability of missing items will generally be lower than the marginal probability of nonmissing items).

Fourth, the latent trait model and its interpretation are generally robust to different item structures, while the sum-of-item-responses approach is not. To see this, note that the expected sum of item responses shown in Figure 7 is a nonlinear function of θ_i^*. The curve describing the expected sum of item responses is steepest in that part of the distribution of θ where the marginal probability curves (Figure 4) are, on average, steepest. In general, this will be where the item marginal severities are most densely concentrated. As a result, the relationship between both the observed and expected sums of item responses and θ_i^* will be sensitive to the item structure used. Adding or removing items from the survey or the analysis will therefore alter the interpretation of the metric of the sum of item responses. It will not, however change the interpretation of the metric of θ, since θ will always be expressed in a well-defined interval-scale metric relative to the conditional log-odds of item responses.

Finally, the partial independence item response model is general enough to lend itself to a number of possible applications and extensions. It can, for example, be used to construct measurement models that use information from a combination of dichotomous and ordinal items—a situation common to many social surveys. It could, for example, be used to model a discrete-time "multiple hazard" process—for example, where an individual is exposed to multiple, partially independent events (e.g., conditional on having some medical procedure, a person might be at risk for several conditionally independent postoperative complications).

In this paper, we have extended the methodology of latent trait modeling via item response theory to the case of surveys with filter or "gate" items—items for which the responses determine whether subsequent questions are asked. Such data are common in social surveys but violate the conditional independence assumption of standard item response models. The partial independence assumption we rely on here is weaker than the standard conditional independence assumption,

as it requires conditional independence of item responses only given membership in the risk set for each item. Thus, while the Rasch model, the discrete-time hazard model (including the multilevel discrete-time model), and the continuation ratio model are each special cases of our model, the partial independence item response model is more general than each of these.

REFERENCES

Adalbjarnardottir, S.. 2002. "Adolescent Psychosocial Maturity and Alcohol Use: Quantitative and Qualitative Analysis of Longitudinal Data." *Adolescence* 37(145):18–53.

Adams, R. J., M. R. Wilson, and W. C. Wang. 1997. "The Multidimensional Random Coefficients Multinomial Logit Model." *Applied Psychological Measurement* 21:1–23.

Allison, P. D.. 1982. "Discrete-Time Methods for the Analysis of Event Histories." Pp. 61–98 in *Sociological Methodology*, vol. 13. edited by Samuel Leinhardt. San Francisco, CA: Jossey-Bars.

Armstrong, B. G., and M. Sloan. 1989. "Ordinal Regression Models for Epidemiologic Data." *American Journal of Epidemiology* 129(1):191–204.

Bailey, S. L., R. L. Flewelling, and J. V. Rachal. 1992. "Predicting Continued Use of Marijuana Among Adolescents: The Relative Importance of Drug-Specific and Social Context Factors." *Journal of Health and Social Behavior* 33(March):51–66.

Barber, J. S., S. Murphy, W. G. Axinn, and J. Maples. 2000. "Discrete-Time Multilevel Hazard Analysis." Pp. 201–35 in *Sociological Methodology*, vol. 30, edited by Michael E. Sobel and Mark P. Becker. Boston, MA: Blackwell Publishing.

Barnes, G. M., A. S. Reifman, M. P. Farrell, and B. A. Dintcheff. 2000. "The Effects of Parenting on the Development of Adolescent Alcohol Misuse: A Six-Wave Latent Growth Model." *Journal of Marriage and the Family* 62:175–86.

Bennett, M. E., B. S. McCrady, V. Johnson, and R. J. Pandina. 1999. "Problem Drinking from Young Adulthood to Adulthood: Patterns, Predictors and Outcomes." *Journal of Studies on Alcohol* 60:605–14.

Chassin, L., S. C. Pitts, and J. Prost. 2002. "Binge Drinking Trajectories from Adolescence to Emerging Adulthood in a High-Risk Sample: Predictors and Substance Abuse Outcomes." *Journal of Consulting and Clinical Psychology* 70(1):67–78.

Cheong, Y. F., and S. W. Raudenbush. 2000. "Measurement and Structural Models for Children's Problem Behaviors." *Psychological Methods* 5(4):477–95.

Colder, C. R., and E. Stice. 1998. "A Longitudinal Study of the Interactive Effects of Impulsivity and Anger on Adolescent Problem Behavior." *Journal of Youth and Adolescence* 27(3):255–74.

Cox, C. 1988. "Multinomial Regression Models Based on Continuation Ratios." *Statistics in Medicine* 7(3):435–41.

Duncan, T. E., S. C. Duncan, and H. Hops. 1996. "The Role of Parents and Older Siblings in Predicting Adolescent Substance Use: Modeling Development via Structural Equation Latent Growth Methodology." *Journal of Family Psychology* 10(2):158–72.

Fienberg, S. E. 1980. *The Analysis of Cross-classified Categorical Data*. 2d ed. Cambridge, MA: MIT Press.

Hill, K. G., H. R. White, I.-J. Chung, J. D. Hawkins, and R. F. Catalano. 2000. "Early Adult Outcomes of Adolescent Binge Drinking: Person- and Variable-Centered Analyses of Binge Drinking Trajectories." *Alcoholism: Clinical and Experimental Research* 24(6):892–901.

Holland, P. W., and H. Wainer, eds. 1993. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Hussong, A. M., P. J. Curran, and L. Chassin. 1998. "Pathways of Risk for Accelerated Heavy Alcohol Use Among Adolescent Children of Alcoholic Parents." *Journal of Abnormal Child Psychology* 26(6):453–66.

Khoo, S., and B. Muthen. 2000. "Longitudinal Data on Families: Growth Modeling Alternatives." Pp. 43–78 in *Multivariate Applications in Substance Use Research: New Methods for New Questions*, edited by J. S. Rose, L. Chassin, C. C. Presson, and S. J. Sherman. Mahwah, NJ: Lawrence Erlbaum.

Little, R. J. A., and D. B. Rubin. 2002. *Statistical Analysis with Missing Data*. 2nd ed. New York: Wiley.

Lord, F. M., and M. R. Novick. 1968. *Statistical Theory of Mental Test Scores*. Reading, MA: Addison-Wesley.

Pinheiro, P. C., and D. M. Bates. 1995. "Approximations to the Log-Likelihood Function in the Nonlinear Mixed Effects Model." *Journal of Computational and Graphical Statistics* 4:12–35.

Raudenbush, S. W., A. S. Bryk, Y. F. Cheong, and R. Congdon. 2005. *Hierarchical Linear and Nonlinear Modeling, Version 6: Users' Guide and Software Program*. Chicago: Scientific Software International.

Raudenbush, S. W., C. Johnson, and R. J. Sampson. 2003. "A Multivariate, Multi level Rasch Model for Self-Reported Criminal Behavior." Pp. 169–211 in *Sociological Methodology*, vol. 33, edited by Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.

Raudenbush, S. W., M.-L. Yang, and M. Yosef. 2000. "Maximum Likelihood for Generalized Linear Models with Nested Random Effects via High-Order, Multi-variate Laplace Approximation." *Journal of Computational and Graphical Statistics* 9(1):141–57.

Reardon, S. F., R. T. Brennan, and S. L. Buka. 2002. "Estimating Multilevel Discrete-Time Hazard Models Using Cross-Sectional Data: Neighborhood Effects on the Onset of Adolescent Cigarette Use." *Multivariate Behavioral Research* 37(3):297–330.

Reckase, M. 1985. "The Difficulty of Items that Measure More Than One Ability." *Applied Psychological Measurement* 9:401–12.

Sampson, R. J., J. D. Morenoff, and S. W. Raudenbush. 2005. "Social Anatomy of

Racial and Ethnic Disparities in Violence." *American Journal of Public Health* 95(2):224–32.

Scheier, L. M., G. J. Botvin, K. W. Griffin, and T. Diaz. 2000. "Dynamic Growth Models of Self-Esteem and Adolescent Alcohol Use." *Journal of Early Adolescence* 20(2):178–209.

Schulenberg, J., K. N. Wadsworth, P. M. O'Malley, J. G. Bachman, and L. D. Johnston. 1996. "Adolescent Risk Factors for Binge Drinking During the Transition to Young Adulthood: Variable- and Pattern-Centered Approaches to Change." *Developmental Psychology* 32(4):659–74.

Silberg, J., M. Rutter, B. D'Onofrio, and L. Eaves. 2003. "Genetic and Environmental Risk Factors in Adolescent Substance Use." *Journal of Child Psychology and Psychiatry* 44(5):664–76.

Singer, J. D., and J. B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Wills, T. A., and S. D. Cleary. 1999. "Peer and Adolescent Substance Use Among 6th–9th Graders: Latent Growth Analyses of Influence Versus Selection Mechanisms." *Health Psychology* 18(5):453–63.