

# Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions

TULSHI D. SAHA, S. PATRICIA CHOU AND BRIDGET F. GRANT\*

*Laboratory of Biometry and Epidemiology, Division of Intramural Clinical and Biological Research, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA*

## ABSTRACT

**Background.** Item response theory (IRT) was used to determine whether the DSM-IV diagnostic criteria for alcohol abuse and dependence are arrayed along a continuum of severity.

**Method.** Data came from a large nationally representative sample of the US population, 18 years and older. A two-parameter logistic IRT model was used to determine the severity and discrimination of each DSM-IV criterion. Differential criterion functioning (DCF) was also assessed across subgroups of the population defined by sex, age and race-ethnicity.

**Results.** All DSM-IV alcohol abuse and dependence criteria, except alcohol-related legal problems, formed a continuum of alcohol use disorder severity. Abuse and dependence criteria did not consistently tap the mildest or more severe end of the continuum respectively, and several criteria were identified as potentially redundant. The drinking in larger amounts or for longer than intended dependence criterion had the greatest discrimination and lowest severity than any other criterion. Although several criteria were found to function differentially between subgroups defined in terms of sex and age, there was evidence that the generalizability and validity of the criterion forming the continuum remained intact at the test score level.

**Conclusions.** DSM-IV diagnostic criteria for alcohol abuse and dependence form a continuum of severity, calling into question the abuse–dependence distinction in the DSM-IV and the interpretation of abuse as a milder disorder than dependence. The criteria tapped the more severe end of the alcohol use disorder continuum, highlighting the need to identify other criteria capturing the mild to intermediate range of the severity. The drinking larger amounts or longer than intended dependence criterion may be a bridging criterion between drinking patterns that incur risk of alcohol use disorder at the milder end of the continuum, with tolerance, withdrawal, impaired control and serious social and occupational dysfunction at the more severe end of the alcohol use disorder continuum. Future IRT and other dimensional analyses hold great promise in informing revisions to categorical classifications and constructing new dimensional classifications of alcohol use disorders based on the DSM and the ICD.

\* Address for correspondence: Professor Bridget F. Grant, Laboratory of Epidemiology and Biometry, Room 3077, Division of Intramural Clinical and Biological Research, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, 5635 Fishers Lane, MS 9304, Bethesda, MD 20892-9304, USA.

(Email: bgrant@willco.niaaa.nih.gov)

The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of any of the sponsoring agencies or the US government.

## INTRODUCTION

Historically, much of the research in the alcohol field has conceived of alcohol problems as defining distinct typologies (Jellinek, 1960; Schuckit, 1985; Cloninger, 1987). Initially, cluster analytic techniques were used to identify

distinct subtypes of alcoholics (Babor *et al.* 1992; Schulenberg *et al.* 1996). However, these techniques were not without problems, often producing different solutions for the same data and difficulties in interpreting the number of clusters for a given data set (Blashfield, 1984).

In view of these limitations, researchers have recently taken an exploratory approach to identifying subtypes of individuals with the same alcohol problem profiles using a more rigorous statistical approach, latent class analysis (LCA) (Bucholz *et al.* 1996, 2000; Kendler *et al.* 1998; Nelson *et al.* 1999; Chung & Martin, 2001). Like the earlier cluster analytic techniques, the LCA approach also proved problematic. In LCA, which uses categorical latent variables, the latent classes ignore possible within-class heterogeneity, such as individual differences in severity. For this reason, among others, subtypes of alcoholism emerging from LCA studies were not distinguished by unique profiles or classes but rather by their placement along a continuum of severity (Heath *et al.* 1994).

Although the search for discrete subtypes or distinct profiles of alcohol problems has yielded evidence for a continuum of severity, very little research has applied modern dimensional psychometric methods to inform conceptualizations of alcohol problems. Using factor analytic approaches, both exploratory and confirmatory, only a few studies have explored the factor structure underlying alcohol problems, and most of them examined the dimensionality of diagnostic criteria or symptom items underlying the criteria of the DSM-III-R (APA, 1987). The results of most of these studies were generally consistent in showing support for similar two-factor models of alcohol problem criteria that bore a similar, but imperfect, relationship to the prevailing DSM-III-R diagnostic classification (Muthén *et al.* 1993*a,b*; Muthén, 1995), while others found evidence for a single dimension based on DSM-IV (APA, 1994) criteria (Nelson *et al.* 1999). Although differences between DSM-III-R and DSM-IV criteria may account for differences in the number of factors observed, all of these studies demonstrated invariance of their empirically derived factor structures to various degrees across sociodemographic subgroups of the population. However, factor structure equivalence does not imply scalar

equivalence, which holds when scores represent the same levels of a construct across diverse populations (van de Vijver & Leung, 1997), as criterion mean scores are typically ignored in factor analysis.

In an attempt to overcome this limitation, researchers from numerous perspectives have begun to gravitate towards item response theory (IRT) (Rasch, 1960; Birnbaum, 1968; Lord, 1980), which provides an appealing framework for studying scalar equivalence (Embretson & Reise, 2000). Unlike factor analytic techniques, IRT can characterize differences in criterion functioning in a way that does not depend on differences in the distribution of the latent construct (i.e. alcohol use disorder) across groups being compared (Bolt *et al.* 2004). To date, only a few studies in the alcohol field (Kahler *et al.* 2003*a,b*; Krueger *et al.* 2004; Langenbucher *et al.* 2004) have applied IRT to examine the possibility of an alcohol problem continuum in which multiple alcohol problems map to a broad dimension of severity. However, these studies were limited to 1348 fathers who participated in the Minnesota Twin Family Study, 372 adult patients, 194 adults mandated to a domestic violence intervention program, and 166 alcohol-dependent patients, samples that are not representative of all individuals with alcohol problems. To determine whether alcohol problems measure a unitary dimension of severity, large representative general population samples are needed. More importantly, the sample sizes of these studies were far too small to capitalize on IRT methodology to examine whether each criterion functioned differently among subtypes of the general population defined in terms of sex, age and race-ethnicity in a manner independent of the distribution of the construct across these groups. The determination of differential criterion functioning (DCF) is important because criteria that exhibit DCF are of questionable validity and may represent bias in the assessment of alcohol use disorder. Moreover, the presence of DCF indicates that the odds of endorsing a particular criterion is not invariant across sociodemographic characteristics or other unique experiences of the individual.

Accordingly, the purpose of the present study was to determine whether the abuse and dependence criteria of the DSM-IV (APA, 1994)

defined an alcohol use disorder continuum using a large ( $n=43\,093$ ) nationally representative sample of the US population, the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) of the National Institute on Alcohol Abuse and Alcoholism (NIAAA). IRT methodology was used: (1) to examine the ability of each DSM-IV alcohol abuse and dependence criterion to discriminate between individuals across the alcohol use disorder continuum; (2) to determine the differential severity of DSM-IV alcohol abuse and dependence criteria; and (3) to identify a subset of DSM-IV criteria that conveyed the most information along the entire breadth of the continuum. In addition, the large sample size of the NESARC allowed for the examination of DCF across important sex, age and race-ethnic subgroups of the population.

## METHOD

### Sample

The 2001–2002 NESARC is a representative sample of the USA conducted by NIAAA, as described elsewhere (Grant *et al.* 2003, 2004). The NESARC target population was the civilian non-institutionalized population residing in households and group quarters, 18 years and older. Face-to-face interviews were conducted with 43 093 respondents, with a response rate of 81%. Blacks, Hispanics and young adults (ages 18–24 years) were oversampled, with data adjusted for oversampling and household- and person-level non-response. The weighted data were then adjusted to represent the US civilian population based on the 2000 census. The sample for this study was restricted to 22 526 respondents, classified as current drinkers (i.e. those who drank at least 12 drinks of alcohol in the year preceding the interview and ever drank 5+ drinks on at least one occasion). Current as opposed to lifetime data were used to avoid the effects of recall bias.

### DSM-IV alcohol abuse and dependence criteria

The NIAAA Alcohol Use and Disorders and Associated Disabilities Interview Schedule – DSM-IV version was designed to generate DSM-IV diagnoses of alcohol abuse and alcohol dependence. These diagnoses were based on 36 symptom items that formed the basis of

the four criteria for alcohol abuse and seven criteria for alcohol dependence. The symptom items underlying the criteria have been described earlier and appear on the NIAAA website (Grant *et al.* 2003, in press). Following DSM-IV, alcohol abuse criteria included recurrent drinking resulting in one of the following during the 12 months prior to the interview: (1) use in hazardous situations; (2) failure to fulfill major role obligations at work/school/home; (3) legal problems related to drinking; and (4) social or interpersonal problems. DSM-IV dependence criteria included: (1) tolerance; (2) withdrawal symptoms or withdrawal relief/avoidance; (3) drinking larger amounts or for longer periods than intended; (4) persistent desire or unsuccessful efforts to cut down or control drinking; (5) a great deal of time spent in activities to obtain alcohol to drink, or to recover from its effects; (6) giving up or reducing important social, occupational or recreational activities in favor of drinking; and (7) continued drinking despite knowledge of a physical or psychological problem caused or exacerbated by drinking.

As reported in detail elsewhere, reliability and validity were good to excellent for alcohol use disorder diagnoses (Grant *et al.* 1995, 2003, 2004; Chatterji *et al.* 1997; Cottler *et al.* 1997; Hasin *et al.* 1997, 2003; Pull *et al.* 1997; Ustun *et al.* 1997; Vrašti *et al.* 1997; Canino *et al.* 1999; Hasin & Paykin, 1999; Nelson *et al.* 1999; Compton *et al.* 2004). Reliability (Chatterji *et al.* 1997) and validity (Cottler *et al.* 1997) for alcohol abuse and dependence criteria were fair to good, as assessed by clinical reappraisal studies conducted by psychiatrists using a semi-structured diagnostic interview. Intra-class correlations of alcohol abuse and dependence criteria derived from a test–retest study of the general population (Grant *et al.* 1995, 2003) were excellent (intra-class correlations = 0.71–0.75), as were the  $\kappa$  values (0.61–0.74).

### Statistical analysis

From several logistic parameterizations, we selected the two-parameter logistic model (Birnbau, 1968; Lord & Novick, 1968), an extension of the Rasch or one-parameter logistic model (Rasch, 1960), to define the relationship between the observed responses to the criteria and the underlying unobserved latent trait or

construct (alcohol use disorder severity):

$$p_{ij}(x_i = 1|\theta_j) = \frac{\exp(a_i(\theta_j - b_i))}{1 + \exp(a_i(\theta_j - b_i))}$$

where the probability  $p$  that a person  $j$  with ability (underlying trait)  $\theta$  will endorse a criterion  $x_i$ , where  $b_i$  is the severity parameter for criterion  $i$  and  $a_i$  is the discrimination parameter for criterion  $i$ . As  $p$  in the equation is the probability of endorsing a criterion,  $1 - p$  is the probability of not endorsing the criterion. This IRT model, generated using the BILOG-MG statistical program (SSI, 2003), yields marginal maximum likelihood estimates (Bock & Aitkin, 1981; Harwell *et al.* 1988) of two parameters: the  $b$  (threshold) parameter and the  $a$  (discrimination) parameter. The  $a$  parameter measures the ability of a criterion to discriminate people who are higher on the continuum and those who are lower on the continuum. This parameter describes how strongly the criterion is related to the underlying trait or construct. The larger the  $a$  parameter (i.e. the slope at its steepest point), the greater the discrimination of a criterion. The  $b$  parameter measures the severity of a criterion; criteria with high thresholds are endorsed less frequently and are more severe.

The  $a$  and  $b$  parameters were plotted graphically as criterion response curves (CRCs). In these plots the  $b$  parameter represents the criterion's location along the latent continuum (located on the horizontal axis). The  $b$  parameter (severity) is the point on the latent continuum where there is a 50% chance of the criterion being present. The  $b$  parameter shifts the CRC from left to right as the criterion becomes more severe. The  $a$  or discrimination parameter indicates how steep the slope of the CRC is at its steepest point.

We also used BILOG-MG (SSI, 2003) to transform the CRCs into criterion information functions (CIFs). The CIFs depict where along the alcohol use disorder continuum each abuse and dependence criterion was conveying the most information. It provides a visual representation of the information value of each criterion. Alcohol use disorder severity is plotted on the  $x$ -axis and the amount of information is plotted on the  $y$ -axis. The CIF of an individual criterion is determined by the  $a$  and  $b$  parameters of its

CRC. The height of the peak of the CIF curve reflects its discrimination parameter ( $a$ ). The higher the curve, the greater the information and discrimination of the criterion. The location of the curve corresponds to its severity parameter ( $b$ ).

In addition to constructing CIFs for each criterion, we constructed an aggregate CIF (ACIF) that graphically depicts the information value of the criteria as a collective or in the aggregate. The ACIF is the reciprocal error variance in an efficient estimate of the latent trait and measures the contribution of each criterion to the reduction of error of measurement. We also constructed a criterion standard error of measurement (S.E.M.) curve to assess the magnitude of error across levels of severity of alcohol use disorder. The S.E.M. curve is the inverse of the square root of the ACIF across all levels of severity.

To determine whether any of the abuse and dependence criteria displayed DCF, we statistically compared  $a$  and  $b$  parameters for each criterion across groups defined by sex (males as the referent category), age (18–24 as the referent category, 25–44 and 45+ years) and race-ethnicity (White as the referent category, Black and Hispanic) using IRT methodology. These DCF analyses were conducted using the PARSCALE program (SSI, 2003), which simultaneously estimated the empirical posterior distribution along with the slope and threshold parameters for each group and computed the  $\chi^2$  test statistics for significance testing among groups. Differences in a criterion's discrimination parameter  $a$  between groups indicate the degree to which a criterion is related to the underlying trait differences between groups, or alternatively that reliability of the criterion varies by group. DCFs related to difference in a criterion's severity parameter  $b$  between groups (e.g. men and women) suggests that unequal levels of the trait are necessary to endorse the criterion.

Criteria that demonstrate DCF need not reflect bias or variance across subgroups if the DCF occurs in opposing directions (e.g. some criteria result in greater discrimination or severity among men while others demonstrate the opposite effect) (Cooke *et al.* 2001; Bolt *et al.* 2004). Whether criteria demonstrating significant DCF do in fact reflect invariance across subgroups can be determined if the observed

Table 1. *Factor analyses*

| Alcohol abuse (A) and dependence (D) criterion | Prevalence (%) | Two-factor model |              | One-factor model  |  |
|--|----------------|------------------|--------------|---|--|
|  |                | Factor 1         | Factor 2     | Factor loadings using all abuse and dependence criteria | Factor loadings eliminating legal problems criterion |
| Tolerance (D)                                  | 8.60           | <b>0.698</b>     | 0.313        | 0.765   | 0.766  |
| Withdrawal (D)                                 | 8.20           | <b>0.805</b>     | 0.221        | 0.810   | 0.822  |
| Larger/Longer (D)                              | 12.90          | <b>0.773</b>     | 0.302        | 0.824   | 0.829  |
| Quit/Control (D)                               | 11.60          | <b>0.660</b>     | 0.292        | 0.721   | 0.721  |
| Time spent (D)                                 | 3.02           | <b>0.816</b>     | 0.344        | 0.883   | 0.885  |
| Activities given up (D)                        | 1.10           | <b>0.836</b>     | 0.326        | 0.892   | 0.896  |
| Physical/Psychological problems (D)            | 4.90           | <b>0.811</b>     | 0.360        | 0.887   | 0.888  |
| Neglect roles (A)                              | 1.30           | <b>0.827</b>     | 0.366        | 0.903   | 0.905  |
| Hazardous use (A)                              | 10.70          | <b>0.562</b>     | 0.466        | 0.716   | 0.701  |
| Legal problems (A)                             | 1.10           | 0.303            | <b>0.953</b> | 0.674   | —  |
| Social/Interpersonal problems (A)              | 2.90           | <b>0.723</b>     | 0.488        | 0.871   | 0.859  |

DCFs cancel out at the total test (scale) score level. To accomplish this, we plotted the expected raw scores by the severity of the alcohol use disorder continuum for age, sex and race-ethnic groups – plots referred to as the test response curves (TRCs). If the TRCs for subgroups (e.g. for men and women) do not substantially differ we can conclude that the significant criterion-level DCFs (if found) cancel out when considered at the total scale level and that for any latent trait value, men and women have identical expected raw scores. If, however, the TRCs do differ substantially between subgroups defined by sex, age or race-ethnicity, individual criteria demonstrating DCF are biased, lacking invariance across important subgroups of the population, and should be eliminated.

Finally, as a preliminary to the above-referenced IRT analyses, we conducted an exploratory factor analysis of tetrachoric correlations using Mplus software (Muthén & Muthén, 2004) to test the ‘essential’ assumption of unidimensionality in the IRT. Although Lord (1952) initially emphasized the primacy of the assumption of trait unidimensionality, and more recent research has supported this position (Ansley & Forsyth, 1985), other research (Reckase, 1979; Drasgow & Parsons, 1983; Harrison, 1986; Hambleton, 1989; Drasgow & Hulin, 1990) has shown that, with sufficient sample sizes and the dominance of one dimension, IRT parameter estimates are stable and accurately represent the underlying data (Farmer *et al.* 2001). However, the *a* and *b* parameters of the IRT model are most interpretable when

the criteria reflect a single unitary dimension (Stout, 1987; Downing, 2003; Bolt *et al.* 2004; Krueger *et al.* 2004). In factor analysis, unidimensionality is established by demonstrating that a one-factor model provides the most parsimonious fit to the data.

## RESULTS

### Factor analysis and unidimensionality

Prior to estimating the parameters of the two-parameter logistic model, we performed an exploratory factor analysis using Varimax rotation to assess unidimensionality of the abuse and dependence criteria (Table 1). The criterion legal problems had the lowest loading in the one-factor model (0.674). After fitting a two-factor model, the legal problems criterion loading on factor 1 diminished (0.303) and its corresponding loading on factor 2 became substantial (0.953). This suggested that the criterion legal problems be dropped to improve the unidimensionality of the criteria and to develop more robust IRT parameters. Further analyses also showed that once the legal problems criterion was dropped from the one-factor model, the resultant root mean square residual decreased from 0.038 to 0.028 and the ratio of eigenvalue 1 to eigenvalue 2 increased from 10.97 to 12.72, both indications of improved fit. We also estimated the two-parameter IRT model with and without the legal problems criterion (data not shown). The IRT model without the legal problems criterion provided a significantly better fit to the data than the model

Table 2. Criterion response model parameters

| Alcohol abuse (A) and dependence (D) criterion | Discrimination ( <i>a</i> ) estimate (s.e.) | Severity ( <i>b</i> ) estimate (s.e.) |
|--|---|---------------------------------------|
| Tolerance (D)                                  | 2.04 (0.05)                                 | 1.80 (0.90)                           |
| Withdrawal (D)                                 | 2.27 (0.06)                                 | 1.76 (0.02)                           |
| Larger/Longer (D)                              | 2.54 (0.07)                                 | 1.34 (0.02)                           |
| Quit/Control (D)                               | 1.76 (0.04)                                 | 1.70 (0.02)                           |
| Time spent (D)                                 | 2.67 (0.10)                                 | 2.29 (0.03)                           |
| Activities given up (D)                        | 2.34 (0.11)                                 | 3.02 (0.92)                           |
| Physical/Psychological problems (D)            | 2.57 (0.08)                                 | 2.04 (0.03)                           |
| Neglect roles (A)                              | 2.47 (0.12)                                 | 2.85 (0.05)                           |
| Hazardous use (A)                              | 1.74 (0.04)                                 | 1.72 (0.03)                           |
| Social/Interpersonal problems (A)              | 2.34 (0.08)                                 | 2.43 (0.04)                           |

including the legal problems criterion; the Bayesian information criterion (BIC) for the model with the legal problems criterion was 82792.38 relative to the significantly lower BIC (80707.03) associated with the model without the legal problems criterion. Accordingly, the legal problems criterion was eliminated in all subsequent IRT analyses.

#### IRT model parameters and criterion response curves (CRCs)

The IRT model parameters and associated CRCs for each abuse and dependence criterion are shown in Table 2 and Fig. 1. The magnitude of the discrimination estimates (*a* parameter) determined the extent to which each criterion was able to discriminate the individuals along the underlying alcohol use disorder continuum. Discrimination was greatest for the time spent, physical/psychological problems and larger/longer criteria, whereas discrimination for the quit/control and hazardous use criteria were among the lowest. With respect to the *b* parameter, two criteria demonstrated the greatest severity (activities given up and neglect of roles), while the larger/longer criterion was the least severe. As shown in Fig. 1, all other criteria were intermediary in terms of severity.

#### Criterion information functions (CIFs)

To identify criteria that provide the most information across the entire alcohol use disorder continuum, we calculated CIFs for each abuse and dependence criterion. As shown in Fig. 2, different criteria contributed different amounts of information and all were contributing

information within similar ranges of the continuum (i.e. the more severe range). The CIFs with the highest peaks and thus the greatest information value were associated with the physical/psychological problems, time spent, social/interpersonal problems and larger/longer criteria, with the longer/larger criterion demonstrating the greatest informational value across a broader range of the alcohol use disorder continuum.

Fig. 3 shows the empirically observed ACIF for the DSM-IV abuse and dependence criteria, which assesses the higher or more severe end of the continuum better than the lower or less severe end of the continuum. The s.e.m. curve clearly shows less error in measurement at the mid- and higher levels of severity compared with the lower levels of the alcohol use disorder continuum.

#### Differential criterion functioning (DCF)

There were few criteria that exhibited DCF by gender. The quit/control criterion had significantly lower ( $p < 0.05$ ) discrimination among men than women. With respect to the *b* parameter, the withdrawal, neglect roles and larger/longer criteria were more severe among men than women, while the opposite was true for the hazardous use and social/interpersonal problems criteria.

Several abuse and dependence criteria demonstrated DCF among age subgroups of the population. In general, discrimination was significantly lower among 25- to 44-year-olds and respondents  $\geq 45$  years compared with 18- to 24-year-olds for the tolerance and hazardous use criteria and significantly greater among the two oldest age groups relative to 18- to 24-year-olds for the activities given up criterion. Severity associated with the activities given up and neglect of roles criteria was significantly greater among 18- to 24-year-olds compared with the two oldest age groups, while the opposite was true for the tolerance criterion. For the time spent criterion, severity was significantly greater among 25- to 44-year-olds relative to 18- to 24-year-olds. Comparisons between 18- to 24-year-olds and respondents  $\geq 45$  years revealed DCFs in opposite directions: significantly greater severity for the withdrawal and larger/longer criteria and significantly lower severity of the quit/control criterion among the oldest relative to the youngest age group.

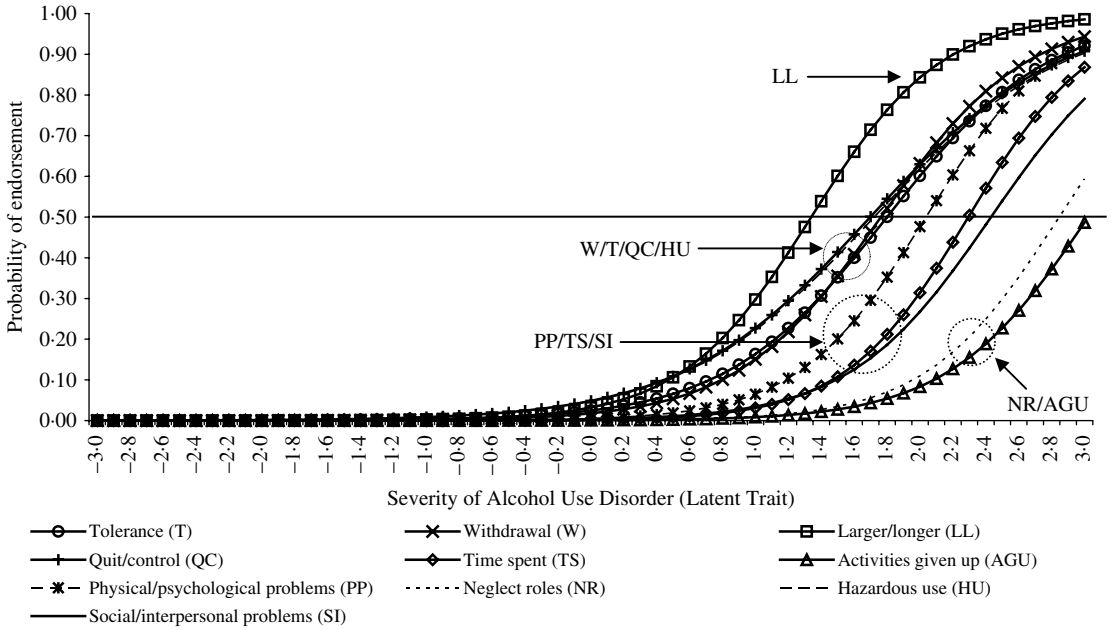


FIG. 1. Criterion response curves.

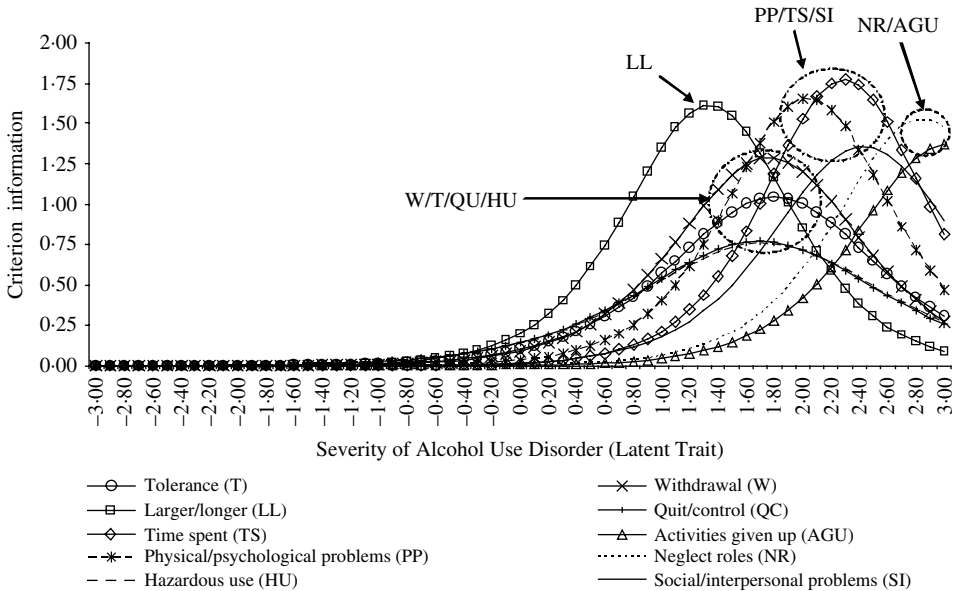


FIG. 2. Criterion information curves.

Little DCF was exhibited by race-ethnicity, and what DCF was demonstrated occurred between Whites and Blacks. Discrimination associated with the tolerance, quit/control and hazardous use criteria was significantly lower

among Blacks than Whites. By contrast, DCF associated with severity occurred in both directions, with the tolerance and quit/control criteria being significantly more severe among Blacks than Whites and the larger/longer and

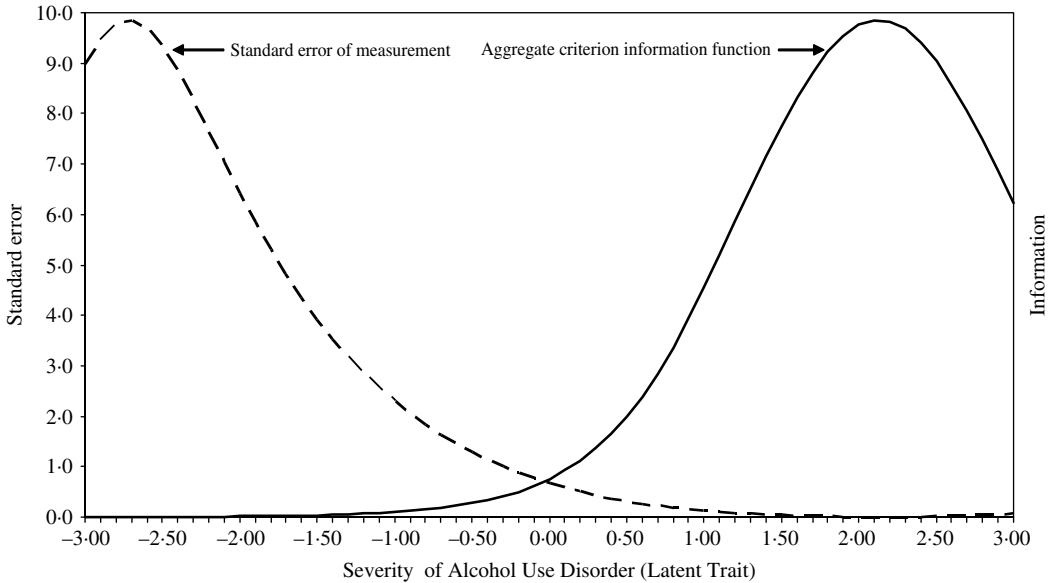


FIG. 3. Aggregate criterion information function and standard error of measurement.

hazardous use criteria being significantly less severe among Blacks than Whites.

Test response curves were essentially identical (i.e. overlapping) for each sex group, each age group and each race-ethnic group. Although some of the alcohol abuse and dependence criteria did show DCF, when considered at the scale level, there was no evidence of DCF between subgroups defined by age, sex and race-ethnicity.

## DISCUSSION

The results of this study showed that DSM-IV alcohol abuse and dependence criteria are arrayed along a continuum of severity. The IRT results questioned the validity of the DSM-IV dependence criteria as distinct from abuse criteria. The criterion information curves identified dependence criteria (e.g. larger/longer, withdrawal, tolerance, quit/control) that were well represented among the mildest criteria, and abuse criteria (e.g. hazardous use) that tapped the more severe range of the continuum. These findings are also at variance with the common interpretation of DSM-IV alcohol abuse as a milder disorder than DSM-IV alcohol dependence.

The legal problems criterion of DSM-IV alcohol abuse was not found to be informative

with regard to the underlying dimension of severity of alcohol use disorder. This criterion demonstrated a weaker relationship to the underlying dimension of alcohol use disorder severity than the other criteria. This finding suggests a re-evaluation of the legal problems criterion as an alcohol use disorder criterion in future revisions of the DSM. Furthermore, the dense clustering of severity and discrimination associated with the withdrawal, tolerance, quit/control and hazardous use criteria suggests that these criteria may be redundant, carrying essentially the same information on case severity. Further research addressing issues of redundancy are needed to possibly reduce the number of alcohol use disorder criteria, with minimal costs to their conceptual integrity and validity.

The criterion information functions for the DSM-IV abuse and dependence criteria were generally peaked at the higher end of the continuum, tapping the more severe range of the alcohol use disorder continuum, suggesting that the DSM-IV criteria assessed in this study would be relevant to the purposes of identifying severe cases in the general population and making fine distinctions between levels of severity among clinical cases for the purpose of matching treatment to illness severity. This finding was not surprising because the criteria



we assessed are part of clinical conceptualizations of alcohol use disorders as defined in the DSM-IV. Nonetheless, the psychometrically informative criteria assessed in this study are relevant for inclusion in future revisions of official classifications, including the International Classification of Diseases (ICD) – Eleventh Edition. However, the scarcity of criteria at the less severe range of the continuum underscores the need to identify other criteria tapping the milder and intermediate ranges of the underlying severity of alcohol use disorder continuum. Such criteria would serve the broader epidemiological purposes of screening, prevention and establishing base rates for alcohol use disorder and the clinical purpose of identifying pre-clinical cases.

The larger/longer dependence criterion uniquely had excellent discrimination and low severity, suggesting that it may be a bridging criterion that links the less severe end of the alcohol use disorder continuum with the more severe end revealed in this study. As the larger/longer criterion represents a high-risk drinking pattern, it is possible that other drinking patterns that incur risk of alcohol use disorder, such as binge drinking or drinking in excess of nationally recommended guidelines, may be good candidates to represent the mild and intermediate ranges of the alcohol use disorder continuum. The results of the present study can serve as a strong starting point for future research that seeks to identify criteria that bridge high-risk drinking patterns and other concepts to tolerance, withdrawal, impaired and compulsive drinking and serious dysfunction in social/interpersonal relationships and role fulfillment.

The neglect of roles abuse criterion, along with the activities given up in favor of drinking criterion, demonstrated the greatest discrimination and severity of all the DSM-IV abuse and dependence criteria. Conceptually, these two criteria suggest a fairly advanced stage of alcohol use disorder – a stage at which drinking becomes so consuming that important social, occupational and recreational activities are given up in favor of drinking, and important life functions at work, school, and home become impaired. In view of these findings, moving the neglect of roles criterion from the DSM-III-R dependence category to the DSM-IV abuse

category seems both conceptually and empirically unsupported. Furthermore, the propensity of both DSM and ICD Substance Use Disorder Working Groups to relegate all criteria that reflect social and interpersonal dysfunction or problems to the ‘milder’ abuse category (Rounsaville *et al.* 1986) is one that can no longer stand up to empirical scrutiny.

This study was the first to examine DCF for alcohol use disorders by age, sex and race-ethnicity. Significant DCF was observed for sex and age and to a lesser extent for sex and race-ethnicity at the criterion level. These results suggest that differences may exist in the expression of certain alcohol use disorder criteria among major subgroups of the population. In general, criteria that demonstrate DCF are of questionable validity and may contribute bias in ordering individuals along the alcohol use disorder continuum. However, the presence of DCF does not necessarily imply criterion bias when the DCFs observed across criteria occur in opposing directions (some criteria demonstrated greater severity and discrimination among men while others were more severe and discriminating among women) and the magnitudes of the observed DCFs are small despite their statistical significance. Indeed, the TRCs showed very little or no differential scale functioning across sex, age and race-ethnic groups. The observed criterion-level DCFs appear not to make a difference when considered at the scale level. Thus, criteria demonstrating DCF in this study are not necessarily biased as they clearly have strong relationships with a common latent variable, alcohol use disorder, within each sex, age and race-ethnic group. The criteria examined in this study appear to work for each age, sex and race-ethnic group as indicators of a common construct, but they work somewhat differently. Future research using other large representative datasets and different criterion measures is needed to replicate the DCF results reported here and to determine the invariance of the alcohol use disorder continuum across important subgroups of the population. Future research should also examine the dimensional properties of criterion symptom items as these and other levels of aggregation may uncover different features related to the dimension.

The major implications of the dimensional nature of alcohol use disorder as defined by

DSM-IV abuse and dependence criteria to psychiatric and clinical assessment are twofold. First, the results of the IRT analyses of this study can be used to inform future revisions of the DSM and ICD. They can empirically refine current categories of alcohol use disorder (e.g. the elimination of the abuse and dependence distinction) and suggest candidate criteria for removal from the classification (e.g. alcohol-related legal problems). Second, discrimination, severity and/or information values associated with specific criteria can be used as weights in the development of a dimensional scale of alcohol use disorder. Alcohol use disorder criteria that are more discriminating and more severe deserve greater weight in determining the extent of individual problems with alcohol. Furthermore, as this dimensional scale is directly related to the criteria of the categorical classification of alcohol use disorder appearing in the DSM-IV, it can form the basis of a second new dimensional classification or axis in further revisions. The availability of an empirically-derived valid and reliable dimensional measure of alcohol use disorder holds great promise in genetic and neuroscience alcohol research, which has heretofore been hindered by categorical representations of alcohol use disorders.

## ACKNOWLEDGMENTS

The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) was conducted and funded by the National Institute on Alcohol Abuse and Alcoholism (NIAAA), with supplemental support from the National Institute on Drug Abuse (NIDA). This research was supported, in part, by the Intramural Program of the National Institute of Health, NIAAA.

## DECLARATION OF INTEREST

None.

## REFERENCES

- Anslley, T. N. & Forsyth, R. A.** (1985). An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. *Applied Psychological Measurement* **9**, 37–48.
- APA** (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised). American Psychiatric Press: Washington, DC.
- APA** (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th edn). American Psychiatric Press: Washington, DC.
- Babor, T. F., Hoffman, M., Delboca, F. A., Hesselbrock, V., Meyer, R. E., Dolinsky, Z. S. & Rounsaville, B.** (1992). Types of alcoholics: evidence for an empirically derived typology based on indication of vulnerability and severity. *Archives of General Psychiatry* **49**, 599–608.
- Birnbaum, A.** (1968). Some latent trait models. In *Statistical Theory of Mental Test Scores* (ed. F. M. Lord and M. R. Norvick), pp. 397–472. Addison-Wesley: Reading, MA.
- Blashfield, R. K.** (1984). *The Classification of Psychopathology: New Kraepelinian and Quantitative Approaches*. Plenum: New York.
- Bock, R. D. & Aitkin, M.** (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika* **46**, 443–445.
- Bolt, D. M., Hare, R. D., Vitale, J. E. & Newman, J. P.** (2004). A multigroup item response theory analysis of the Psychopathy Checklist-revised. *Psychological Assessment* **16**, 155–168.
- Bucholz, K. K., Heath, A. C. & Madden, P. A. F.** (2000). Transitions in drinking in adolescent females: evidence from the Missouri adolescent female twin study. *Alcoholism: Clinical and Experimental Research* **24**, 914–923.
- Bucholz, K. K., Heath, A. C., Reich, T., Hesselbrock, V. M., Kramer, J. R., Nurnberger Jr., J. I. & Schuckit, M. A.** (1996). Can we subtype alcoholism? A latent class analysis of data from relatives of alcoholics in a multicenter family study of alcoholism. *Alcoholism: Clinical and Experimental Research* **20**, 1462–1471.
- Canino, G. J., Bravo, M., Ramirez, R., Febo, V., Fernandez, R. & Hasin, D. S.** (1999). The Spanish Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): reliability and concordance with clinical diagnosis in a Hispanic population. *Journal of Studies on Alcohol* **60**, 790–799.
- Chatterji, S., Saundeeers, J. B., Vrsti, R., Grant, B. F., Hasin, D. & Mager, D.** (1997). Reliability of the alcohol and drug modules of the Alcohol Use Disorder and Associated Disabilities Interview Schedule – Alcohol/Drug-Revised (AUDADIS-ADR): an international comparison. *Drug and Alcohol Dependence* **47**, 171–185.
- Chung, T. & Martin, C. S.** (2001). Classification and course of alcohol problems among adolescents in addictions treatment programs. *Alcoholism: Clinical and Experimental Research* **25**, 1734–1742.
- Cloninger, C. R.** (1987). Neurogenetic adaptive mechanisms in alcoholism. *Science* **236**, 410–416.
- Compton, W. M., Grant, B. F., Collier, J. D., Glantz, M. D. & Stinson, F. S.** (2004). Prevalence of marijuana use disorders in the United States: 1991–1992 and 2001–2002. *Journal of the American Medical Association* **291**, 2114–2121.
- Cooke, D. J., Kosser, D. S. & Michie, C.** (2001). Psychopathy and ethnicity: structural, item and test generalizability of the Psychopathy Checklist – Revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment* **13**, 531–542.
- Cottler, L. B., Grant, B. F., Blaine, J., Mareas, V., Pull, C., Hasin, D. S., Compton, W. M., Rubia-Striepe, M. & Mager, D.** (1997). Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. *Drug and Alcohol Dependence* **47**, 195–205.
- Downing, M.** (2003). Item response theory: applications of modern test theory in medical education. *Medical Education* **37**, 739–745.
- Drasgow, F. & Hulin, C. L.** (1990). Item response theory. In *Handbook of Industrial and Organizational Psychology* (vol. 2, 2nd edn) (ed. M. D. Dunnette and L. M. Hough), pp. 577–636. Consulting Psychologists Press: Palo Alto, CA.
- Drasgow, F. & Parsons, C. K.** (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement* **7**, 189–199.
- Embretson, S. E. & Reise, S. P.** (2000). *Item Response Theory for Psychologists*. Erlbaum: Mahwah, NJ.
- Farmer, W. L., Thompson, R. C., Heil, S. K. R. & Heil, M. C.** (2001). Latent trait theory analysis of changes in item response anchors.

- Technical report no. DSO/FAA/AM-01/4. Office of Aviation Medicine, Federal Aviation Administration: Oklahoma City, OK.
- Grant, B. F., Harford, T. C., Dawson, D. A., Chou, S. P. & Pickering, R. P.** (1995). The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): reliability of alcohol and drug modules in a general population sample. *Drug and Alcohol Dependence* **39**, 37–44.
- Grant, B. F., Harford, T. C., Muthén, B. O., Yi, H., Hasin, D. S. & Stinson, F. S.** (in press). DSM-IV alcohol dependence and abuse: further concurrent validity in the general population. *Drug and Alcohol Dependence*.
- Grant, B. F., Moore, T. C., Shapard, J. & Kaplan, K.** (2003). *Source and Accuracy Statement, Wave 1 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)*. National Institute on Alcohol Abuse and Alcoholism: Bethesda, MD.
- Grant, B. F., Stinson, F. S., Dawson, D. A., Chou, S. P., Dufour, M. C., Compton, W. M., Pickering, R. P. & Kaplan, K.** (2004). Prevalence and co-occurrence of substance use disorders and independent mood and anxiety disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *Archives of General Psychiatry* **61**, 807–816.
- Hambleton, R. K.** (1989). Principles and selected applications of item response theory. In *Educational Measurement* (3rd edn) (ed. R. L. Linn), pp. 147–200. American Council on Education and Macmillan Publishing: New York, NY.
- Harrison, D. A.** (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption. *Journal of Educational Statistics* **11**, 91–115.
- Harwell, M. R., Baker, F. B. & Zwarts, M.** (1988). Item parameter estimation via marginal maximum likelihood and an EM algorithm: a didactic. *Journal of Educational Statistics* **13**, 243–271.
- Hasin, D. S., Muthén, B. & Grant, B. F.** (1997). The dimensionality of DSM-IV alcohol abuse and dependence: factor analysis in a clinical sample. In *Alcoholism: New Research Perspectives* (ed. R. Vrstic), pp. 36–52. Hogrefe and Hubner: Munich, Germany.
- Hasin, D. S. & Paykin, A.** (1999). Alcohol dependence and abuse diagnoses: concurrent validity in a nationally representative sample. *Alcoholism: Clinical and Experimental Research* **23**, 144–150.
- Hasin, D. S., Schuckit, M. A., Martin, C. S., Grant, B. F., Buchholz, K. K. & Helzer, J. E.** (2003). The validity of DSM-IV alcohol dependence: what do we know and what do we need to know? *Alcoholism: Clinical and Experimental Research* **27**, 244–252.
- Heath, A. C., Buchholz, K. K., Sluske, W. S., Madden, P. A. F., Dwindle, S. H. & Dunne, M. P.** (1994). The assessment of alcoholism in surveys of the general community: what are we measuring? Some insights from the Australian Twin Panel Interview Survey. *International Review of Psychiatry* **6**, 295–307.
- Jellinek, F. M.** (1960). *The Disease Concept of Alcoholism*. College and University Press: New Haven, CT.
- Kahler, C. W., Strong, D. R., Hayaki, J., Ramsey, S. E. & Brown, R. A.** (2003a). An item response analysis of the alcohol dependence scale in treatment-seeking alcoholics. *Journal of Studies on Alcohol* **64**, 127–136.
- Kahler, C. W., Strong, D. R., Stuart, G. L., Moorek, T. M. & Ramsey, S. E.** (2003b). Item functioning of the alcohol dependence scale in a high risk sample. *Drug and Alcohol Dependence* **72**, 183–192.
- Kendler, K. S., Karikowski, L. M., Prescott, C. A. & Pederson, N. L.** (1998). Latent class analysis of temperance board registration in Swedish male-male twin pairs born 1902–1949: searching for subtypes of alcoholism. *Psychological Medicine* **28**, 803–813.
- Krueger, R. F., Nicol, P. E., Hicks, B. M., Markon, K. E., Patrick, C. J., Iacono, W. G. & Mague, M.** (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment* **16**, 107–119.
- Langenbucher, J. W., Labouvie, E., Martin, C. S., Sanjuan, P. M., Bavy, L. & Kirisci, L.** (2004). An application of item response theory analysis to alcohol, cannabis and cocaine criteria in DSM-IV. *Journal of Abnormal Psychology* **113**, 72–80.
- Lord, F. M.** (1952). *A Theory of Test Scores*. Psychometric Monographs, No. 7. University of Chicago: Chicago, IL.
- Lord, F. M.** (1980). *Applications of Item Response Theory to Practical Testing Problems*. Erlbaum: Hillsdale, NJ.
- Lord, F. M. & Novick, M. R.** (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley: Reading, MA.
- Muthén, B. O.** (1995). Factor analysis of alcohol abuse and dependence symptom items in the 1988 National Health Interview Survey. *Addiction* **90**, 637–645.
- Muthén, B. O., Grant, B. & Hasin, D.** (1993a). The dimensionality of alcohol abuse and dependence: factor analysis of DSM-III-R and proposed DSM-IV criteria in the 1988 National Health Interview Survey. *Addiction* **88**, 1079–1090.
- Muthén, B. O., Hasin, D. & Wisnicki, K. S.** (1993b). Factor analysis of ICD-10 symptom items in the 1988 National Health Interview Survey on Alcohol Dependence. *Addiction* **88**, 1071–1077.
- Muthén, B. O. & Muthén, L. K.** (2004). *Mplus: Statistical Analysis with Latent Variables (Version 3.01)*. Los Angeles, CA: Muthén & Muthén Inc.
- Nelson, C. B., Rehm, J., Ustun, T. B., Grant, B. F. & Chatterji, S.** (1999). Factor structure for DSM-IV substance disorder criteria endorsed by cannabis, cocaine and opioid users: results of the World Health Organization Reliability and Validity Study. *Addiction* **94**, 843–855.
- Pull, C. B., Saunders, J. B., Mavreas, V., Cottler, L. B., Grant, B. F., Hasin, D. S. K., Blaine, J., Mager, D. & Ustun, B. T.** (1997). Concordance between ICD-10 alcohol and drug use disorder criteria and diagnoses as measured by the AUDADIS-ADR, CIDI and SCAN: results of a cross-national study. *Drug and Alcohol Dependence* **47**, 207–216.
- Rasch, G.** (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Danish Institute for Educational Research: Copenhagen, Sweden.
- Reckase, M. D.** (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics* **4**, 207–230.
- Rounsaville, B. J., Spitzer, R. L. & Williams, J. B.** (1986). Proposed changes in DSM-III-R substance use disorders: description and rationale. *American Journal of Psychiatry* **143**, 463–468.
- Schuckit, M. A.** (1985). The clinical implications of primary diagnostic groups among alcoholics. *Archives of General Psychiatry* **42**, 191–195.
- Schulenberg, J., O'Malley, P. M., Bachman, J. G., Wadsworth, K. N. & Johnston, L. D.** (1996). Getting drunk and growing up: trajectories of frequent binge drinking during the transition to young adulthood. *Journal of Studies on Alcohol* **57**, 289–304.
- SSI** (2003). *Item Response Theory (IRT) from SSI* (ed. M. Du Toit). Scientific Software International: Lincolnwood, IL.
- Stout, W.** (1987). A new item response theory modeling approach with application of unidimensional assessment and ability estimation. *Psychometrika* **55**, 293–326.
- Ustun, B., Compton, W. M., Mager, D., Babor, T., Baiyewu, O., Chatterji, S., Cottler, L. B., Cogus, A., Mavreas, V., Peters, L., Pull, D., Saunders, J., Smeets, R., Stipek, M. R., Vrstic, R., Hasin, D., Room, R., Van De Brink, W., Reiger, D., Baline, J., Grant, B. F. & Sartorius, N.** (1997). WHO study on the reliability and validity of the alcohol and drug use disorder instruments: overview of methods and results. *Drug and Alcohol Dependence* **47**, 161–169.
- van de Vijver, F. J. R. & Leung, K.** (1997). *Methods and Data Analysis for Cross-Cultural Research*. Sage Publications: Newbury Park, CA.
- Vrstic, R., Grant, B. F., Chatterji, S., Ustun, B. T., Mager, D., Olteanu, I. & Nodoi, M.** (1997). The reliability of the Romanian version of the alcohol module of the WHO Alcohol Use Disorder and Associated Disabilities Interview Schedule – Alcohol/Drug – Revised (AUDADIS-ADR) *European Addiction Research* **4**, 144–149.