



Published in final edited form as:

*Drug Alcohol Depend.* 2009 January 1; 99(1-3): 296–308. doi:10.1016/j.drugalcdep.2008.09.001.

## Linking Measures of Adolescent Nicotine Dependence to a Common Latent Continuum

**David R. Strong**

*Alpert Medical School of Brown University and Butler Hospital*

**Christopher W. Kahler and Suzanne M. Colby**

*Center for Alcohol and Addiction Studies and Brown University*

**Pamela C. Griesler and Denise Kandel**

*Columbia University and New York State Psychiatric Institute*

### Abstract

Using the theoretical model of Nicotine Dependence (ND) operationalized within the Diagnostic and Statistical Manual of Mental Disorder, fourth Edition (DSM-IV: American Psychiatric Association, 1994) as a frame of reference, we used methods based in item response theory to link alternative instruments assessing adolescent nicotine dependence severity to a common latent continuum. A multi-ethnic cohort of 6<sup>th</sup>–10<sup>th</sup> graders selected from the Chicago Public Schools (CPS) completed five household interviews over two years. Youth who reported at least some cigarette use in the last 30 days prior to the interviews at waves W3–W5 completed measures of DSM-IV ND, the Modified Fagerstrom Tolerance Questionnaire (mFTQ: Prokhorov et al., 1998) and the Nicotine Dependence Syndrome Scale (NDSS: Shiffman et al., 2004), yielding samples of 253, 241, and 296 respondents at W3–W5, respectively. Confirmatory factor analysis supported a primary dimension of ND. Each instrument's items had complementary and stable relationships to ND across multiple waves of assessment. By aligning symptoms along a common latent ND continuum, we evaluated the consistency of symptoms from different instruments that target similar content. Further, these methods allowed for the examination of the DSM-IV as a continuous index of ND, evaluation of the degree of heterogeneity in levels of ND within groups above and below diagnostic thresholds, and the utility of using the pattern or particular DSM-IV symptoms that led to each score in further differentiating levels of ND. Finally, we examined concurrent validity of the ND continuum and levels of current of smoking at each wave of assessment.

### Keywords

Nicotine Dependence; Item Response Theory; Adolescent; Smoking

---

Corresponding Author: David R. Strong, Ph.D., Associate Professor (Research), Butler Hospital, The Warren Alpert Medical School of Brown University, 345 Blackstone Blvd. Providence, RI, 02906, phone:(401) 455-6294, fax:(401) 455-6424, email: david\_strong@brown.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## 1. Introduction

Clinical and observational longitudinal studies often rely on indices of nicotine dependence when attempting to describe and understand smoking progression among adolescents. Although the Diagnostic and Statistical Manual of Mental Disorders IV (American Psychiatric Association, 1994) Nicotine Dependence (ND) syndrome was developed for adults, evaluation of DSM-IV ND has been used as an important point of reference for researchers attempting to understand smoking behavior among adolescents. The DSM-IV is the default standard for making decisions about whether adolescents should be classified as having a clinical diagnosis of ND. Increased efforts are needed to understand the degree to which diagnostic models for ND developed among adults apply to adolescent smokers and whether symptoms of ND can be used further to differentiate reliably a continuum of ND among adolescents who report varying degrees of smoking behavior.

Despite its origins, the DSM-IV provides a natural starting point for understanding smoking behavior among adolescents as the seven criteria for Substance Dependence have been a default standard for deciding whether an individual has a level of ND that warrants a diagnosis. The DSM-IV has a theoretical basis (Edwards and Gross, 1976), has multiple criteria available to operationalize levels of ND, and has a widely accepted threshold for categorizing smokers diagnostically. However, the theoretical framework for dependence was not developed specifically for tobacco, the decisions for determining the presence or absence of individual criteria can be operationalized differently across diagnostic instruments, and alternative thresholds for diagnosis have not been evaluated systematically (Crowley, 2006). Furthermore, because smokers can attain a diagnosis of ND by meeting any three of seven criteria, the diagnosis can be conceptually heterogeneous (Shiffman et al., 2004) and DSM-IV indices may serve primarily as a descriptive indicator of an end state of dependence rather than as a tool for understanding mechanisms leading to increasing levels of dependence (Piper et al., 2004). Despite the need for continued evaluation, existing interviews designed to assess the DSM-IV have been useful in describing adolescent smoking behavior (Kandel et al., 2005, 2007; Strong et al., 2003a, 2007) and have demonstrated good inter-rater reliability and validity for use with adolescents (Crowley et al., 2001; Mikulich et al., 2001; Young et al., 1995, 2002). This broad support positions the DSM-IV as a benchmark for the evaluation of measures of ND for adolescent research.

Other models of ND have been offered in the form of assessment instruments each with a different focus or content. Three widely used measures of ND used in adolescent research on smoking behavior include the Modified Fagerstrom Tolerance Questionnaire (mFTQ; Prokhorov et al., 1998), the Nicotine Dependence Syndrome Scale (NDSS; Shiffman et al., 2004), and the Hooked On Nicotine Checklist (HONC; DiFranza et al., 2000). The mFTQ is an adapted version of a scale originally designed to assess behavioral indicators of physical dependence on nicotine among adult smokers (FTQ; Fagerstrom, 1978) in order to evaluate the outcome of smoking cessation interventions. The conceptual model underlying the mFTQ was not designed as a proxy for the DSM-IV criteria. The FTQ and variants including the Fagerstrom Test for Nicotine Dependence (FTND; Heatherton et al., 1991) scale have a long history of support as valid indicators of smoking behavior and as predictors of difficulty with smoking cessation among adults (Alterman et al., 1999; Campbell et al., 1996; Patten et al., 2001; Westman et al., 1997). As a natural extension of this research, the mFTQ included a modified set of adult symptoms for adolescents to address graded levels of smoking behavior, avoidance of the physiological effects of nicotine deprivation, and perceived difficulty refraining when prevented from smoking. Among adolescents, the mFTQ has been shown to predict continued smoking, quantity and frequency of tobacco use, and shorter periods of abstinence from smoking over time (Cohen et al., 2002; Kandel et al., 2005; Prokhorov et al., 1998; Rojas et al., 1998; Strong et al., 2003a), although studies among light smokers have not

consistently demonstrated relations between mFTQ scores and smoking behavior perhaps in part due to the orientation of the questions to adolescents with more frequent smoking behavior (Etter et al., 1999; Sledjeski et al., 2007; Wellman et al., 2006). The NDSS was based on the same underlying conceptual model of dependence as the DSM-IV (Edwards and Gross, 1976) and includes symptoms assessing five domains: 'Drive' (craving and withdrawal); 'Priority' (behavioral preference of smoking over other reinforcers); Tolerance (reduced sensitivity to smoking effects); 'Continuity' (regularity of smoking patterns); and Stereotypy (other events such as moods or environments do not disturb smoking patterns). Among college students (Sledjeski et al., 2007) and adolescents (Clark et al., 2005), the total score on the NDSS has been most consistently linked to continued smoking, quantity and frequency of tobacco use, and shorter periods of abstinence from smoking over time. Although the mFTQ and NDSS are based on separate theoretical models, both attempt to assess degrees of dependence on nicotine using self-reports of smoking experiences and behavioral symptoms. However, little is known about how much these alternative measures provide unique or overlapping information about the construct of ND assessed by the DSM-IV criteria. A more recent measure is the HONC (DiFranza et al., 2000; Wheeler et al., 2004). This 10-symptom instrument was specifically developed for adolescents to measure loss of autonomy over tobacco use, which is assumed to be the core component of dependence (DiFranza, 2002). The HONC is based on the premise that symptoms of dependence can occur prior to regular heavy smoking and is explained by the neurophysiological action of nicotine (DiFranza and Wellman, 2005). Because the present study does not include data from the HONC, the instrument will not be considered further here.

Attempts to understand the interrelationship of DSM-IV, mFTQ, and NDSS measures of ND have typically used correlations and comparisons of the reliability or incremental validity of these measures in predicting smoking intensity in samples of adolescents (Clark et al., 2005). There is a strong relationship between the NDSS and mFTQ ( $r = 0.60$ ; Clark et al., 2005). While less is known about relationships of the NDSS with DSM-IV indices, the DSM-IV and mFTQ have shown similarly strong relationships in a school sample ( $r = 0.56$ ; Kandel et al., 2005). But these relationships have been more moderate in heavier smoking adolescent populations ( $r = 0.22-0.35$ ; Cohen et al., 2002; Strong et al., 2003b) and among young adults (Breslau and Johnson, 2000). Attempts to examine incremental predictive validity have suggested that although measures based on the FTQ have strong relationships with smoking progression in adolescence, the NDSS total score can provide unique information about smoking progression beyond that explained by other measures (Clark et al., 2005). These studies suggest the possibility of substantial overlap in the measurement of a common construct of ND as well as the potential for increasing reliability of ND assessment among adolescents by combining content from multiple measures.

Recent efforts (Strong et al., 2007) to examine continuous DSM-IV and mFTQ measures of ND in a clinical adolescent sample support the ability to link symptoms to a unidimensional latent construct using methods based on item response theory (Lord, 1980). Results suggested that the DSM-IV and mFTQ symptoms represented overlapping and complementary indicators of the severity of nicotine dependence as indexed by the latent construct. Further, these symptoms collectively were sensitive to observed changes in smoking rates, readiness to make changes in smoking, and confidence in the ability to quit smoking. While findings provided support for this approach to assessing changes in ND symptoms among adolescent smokers, the use of a relatively small sample of psychiatric inpatients with a high proportion of regular daily smokers prevents generalization to non-clinical populations.

The current study was designed to address three primary objectives. First, we assessed the ability to link DSM-IV, mFTQ, and NDSS symptoms in adolescence to a stable unidimensional latent continuum of ND using an item response model and to determine the degree to which

the symptoms from the three scales could be used to index ND in a general population sample of adolescent smokers with more representative levels of adolescent nicotine exposure than psychiatric inpatients. To meet this objective, we examined relationships of symptoms to the latent continuum of ND using data from three repeated waves of assessment. Three waves of data are used as replicates and as validations of the latent continuum; they are not used to examine changes on the severity dimension over time within individuals. Second, given that the DSM-IV was designed specifically to classify adolescents diagnostically, we examined the degree of heterogeneity of responses to the DSM-IV by identifying the most typical patterns of DSM-IV symptoms that led to scores above and below diagnostic thresholds. We then explored whether the pattern of particular DSM-IV symptoms that led to each score revealed information about levels of ND beyond knowing how many symptoms each individual endorsed. Finally, we examined concurrent validity by measuring the association between levels of ND and levels of current of smoking. We expected that increased levels of ND would be related to increased levels of smoking.

## 2. Methods

### 2.1. Data collection

The analyses are based on three waves of interviews with a subsample from a multi-ethnic cohort of 1039 6th–10th graders from the Chicago Public Schools (CPS). Details of the two-stage design implemented to select efficiently the target sample of smokers for follow-up is described in Kandel et al. (2005, 2007). In Phase I (spring 2003), 15,763 students in grades 6–10 were sampled from 43 public schools in the CPS. The sample was designed to provide approximately equal numbers of adolescents among the three major ethnic groups: non-Hispanic white, non-Hispanic African American, and Hispanic. Students were administered a brief questionnaire through eight surveys staggered over 4 months; the completion rate for the survey was 83.1%. In Phase II, on average 9 weeks after each school survey, 1039 (272 white, 343 African American, 424 Hispanic) agreed to participate in the follow-up (84.1% of 1236 targeted youths) consisting of three annual computerized household interviews with youths and parents, each about 90 min long (Waves: W1, W3, W5), and two short bi-annual interviews with youths 6 months after Waves 1 and 3 (Waves: W2, W4). Completion rates at each successive wave were 96% of the original sample ( $N = 996, 999, 1001$ , Waves: W3–W5). The current study included data only from W3, W4, and W5.

Data were collected by the National Opinion Research Center of the University of Chicago (NORC). All procedures for obtaining parental consent and youth assent were approved by the Institutional Review Boards of the New York State Psychiatric Institute and Columbia University.

**2.1.1. Symptoms of nicotine dependence**—Three scales were used to measure dependence: (1) as per the Diagnostic and Statistical Manual of Mental Disorder, fourth Edition (DSM-IV; APA, 1994), (2) the mFTQ; and (3) the NDSS. The DSM-IV and the mFTQ were administered five times at W1–W5; the NDSS was administered three times at W3–W5.

**2.1.1.1 DSM-IV Nicotine Dependence:** The instrument used to measure the seven criteria of DSM-IV nicotine dependence was designed for adolescents and young adults by the Brown University Transdisciplinary Tobacco Use Research Center (TTURC) and evaluated by the Tobacco Etiology Research Network. This instrument has demonstrated reliability (Cronbach Alpha = .75), and predicted continued smoking, quantity and frequency of tobacco use, and shorter periods of abstinence from smoking over time (Sledjeski et al., 2007). The presence of each of the seven DSM-IV dependence criteria were based on one or two items querying tolerance, withdrawal, impaired control, unsuccessful attempts to quit, great deal of time spent

using tobacco, neglect of important activities, and use despite physical or psychological problems. The instrument inquired about experiences in the period since each preceding assessment (average Cronbach Alpha= 0.70).

**2.1.1.2 Modified Fagerstrom Tolerance Questionnaire (mFTQ):** The mFTQ consists of seven symptoms inquiring about quantity of cigarettes smoked, inhaling smoke, time to first cigarette of the day, whether the first cigarette is the most difficult to give up, difficulty refraining from smoking, smoking if ill in bed, and smoking more during the first two hours of the day than at other times. The mFTQ has demonstrated adequate reliability across published studies (Cronbach Alpha = 0.58–0.63). The mFTQ inquired about smoking in the past 30 days prior to each of the 5 waves. Given the low level of smoking observed across these waves of data we dichotomized the quantity smoked symptom at a threshold of >1 cigarette each day (average Cronbach Alpha= 0.61).

**2.1.1.3 Nicotine Dependence Syndrome Scale (NDSS):** The NDSS is a seventeen symptom index that asks teens to rate a series of statements for how true (Coded 1–5 for ‘Not true at all’ to ‘Extremely true’) each statement is for them in the past 30 days. Five separate subscales (drive, priority, continuity, stereotypy, and tolerance) can be scored along with a total score. As the NDSS allows 5 levels of response for each symptom, additional analyses were conducted to evaluate the reliability of these multiple response categories. In the current study, the NDSS had high internal consistency at each wave (average Cronbach Alpha= 0.90).

## 2.2 Analysis plan

Our analyses were designed to establish the degree to which each of the three measures of nicotine dependence provides information about a common continuum of nicotine dependence (ND). Before placing each measure on a common continuum we first established the degree to which each measure could be explained by a primary underlying dimension of ND. Having established that each measure assessed primarily one core construct at each wave, we used confirmatory factor analytic models to evaluate the degree to which the DSM-IV, mFTQ, and NDSS provided unique or overlapping information about the levels of ND observed across these three waves of assessment. With a focus on a single primary dimension of ND, we fit an item response model to evaluate: 1) the relative severity of symptoms from each of the three measures; 2) the ability of each symptom to maintain a stable relationship to levels of ND at each wave and thus serve as a replicable marker for ND at each assessment; and 3) where along the continuum symptoms from these measures provide the most information about levels of ND.

We fit a 2-parameter logistic item response model (IRM) that scales the severity of both symptoms and persons along an underlying latent continuum. The odds of an individual endorsing a given symptom is modeled as a function of the individual’s overall level of ND, the severity of that symptom, and how each symptom discriminates across the ND continuum (Birnbaum, 1968). In the context of this statistical model, symptom “severity” is defined by the point on a latent severity continuum at which the symptom has a 50% likelihood of being endorsed.

## 2.3. Analyses

**2.3.1. Unidimensionality and local independence—**The primary assumption of item response models is that responses to symptom queries are a function of individual variation along a single underlying dimension of nicotine dependence. Secondly, responses to a given symptom should be independent from responses to other symptoms (e.g. locally independent). Statistically, if a unidimensional item response model is applied to multidimensional data, the resulting information about symptoms may be biased (Reckase, 1997). Before combining



measures, we wanted to ensure that each measure had a strong single dimension underlying responses and contained symptoms that could contribute unique rather than overlapping information so as to minimize local dependence. Therefore, before interpreting the fit of the data to an item response model, we first conducted a series of exploratory factor analyses (EFAs) of each measure to estimate the size of the first factor in relation to secondary factors and to ensure strong discrimination among all symptoms (loadings  $> .30$ ).

In a second step, to evaluate the strength of a primary ND factor underlying a combination of all three measures, we conducted a series of confirmatory factor analyses comparing the fit of three alternative models: (a) a single primary dimension (1-factor model); (b) a model that allows three related dimensions defined by the DSM-IV, mFTQ, and NDSS; (c) a hierarchical bifactor model that includes a primary dimension of nicotine dependence and three secondary factors reflecting each measure. By fitting a hierarchical bifactor model, we can evaluate the degree of influence (e.g. local dependence) from secondary factors that may bias estimates of relationships with the primary dimension of ND (Reise et al., 2007). The bifactor model estimates the loadings on a general factor of ND severity after controlling for any influence from the three secondary factors. Loadings from the unidimensional model of ND can be compared to the bifactor model results to assess the impact of any local dependence on the estimates of each symptom's relationship to the primary dimension of ND severity. Our intention in fitting the series of models was not to suggest that the symptoms from the three measures would fit a strict unidimensional model better than a multidimensional model. Instead, we assumed that some degree of variability within measures would remain after controlling for the relationship of symptoms to a primary ND dimension. We present three indices in testing the fit of each of three models: the Comparative Fit Index (CFI: Bentler, 1990), the Tucker Lewis Index (TLI: Bentler and Bonnett, 1980), and the root mean square error of approximation (RMSEA: Steiger, 1990). Cut-offs for model fit have been suggested to be  $CFI \geq 0.96$ ,  $TLI \geq 0.95$ , and  $RMSEA \leq 0.05$  (Yu, 2002). Browne and Cudeck (1993) suggested guidelines for approximate fit values and proposed that RMSEA values of 0.05 or less indicate a 'close fit', values between 0.05–0.09 indicate 'reasonable fit', and values of 0.10 and greater demonstrate 'poor fit.' By fitting the three models, we can observe the degree to which the three factors are interrelated (model b) or may suggest independent processes and the degree to which influences outside of a common ND dimension might bias information from a unidimensional item response model (Reise et al., 2007). Support for proceeding with a unidimensional IRM would include the lack of evidence for significant differences in loadings on a primary dimension after controlling for secondary factors in the hierarchical bifactor model (Reise et al., 2007).

**2.3.2. Fit to the item response model**—To select an item response model we compared two competing models. First we estimated the level of severity of each symptom using a model that required all symptoms to discriminate equally well across levels of ND (e.g., 1-pl). To this model we compared the improvement in fit of a model that also allowed levels of discrimination to be different for each symptom (e.g. 2-pl). Model selection was based on a Likelihood Ratio Test (LRT) that involved subtracting the log-likelihood values ( $ll$ ) for the models being compared. Having selected an appropriate item response model, we then examined the fit of each symptom using a chi-square based index that evaluated the difference between the observed proportion and model-expected proportion of responses across levels of nicotine dependence. Each competing IRM was fit using Marginal Maximum Likelihood Estimation using the package *ltm* (Rizopoulos, 2006) available in R software (R Core, 2007).

**2.3.3. Replication of the model over time**—To assess the stability of the item response model over the three waves of assessment, we evaluated differences in estimates of symptom characteristics obtained from each wave. If the IRM is stable, the estimates of symptom characteristics from each wave should be similar. Replication of the model estimates across

waves of assessment provides support for the construct validity of the ND continuum. We employed a differential item functioning (DIF) approach that allows the use of model-based evaluations that utilize information about the measurement properties of the set of symptoms simultaneously in both groups to generate a posterior distribution of ND severity. Following Thissen et al. (1995), we used a likelihood-ratio test statistic to provide a significance test for the null hypothesis that the symptom parameters did not differ across waves. We employed Version 2.0 of IRTLRDIF (Thissen, 2001) to complete DIF analyses. Analyses are conducted iteratively to establish which symptoms appear to function differently across assessments and which symptoms are relatively DIF-free. Symptoms that are DIF-free can serve as an anchor for evaluating the symptoms suspected of having significant DIF. Separate anchors were established for analyses of W3 vs. W4, W3 vs. W5, and W4 vs. W5 data. Analyses of DIF symptoms proceed by initially constraining both the discrimination and the severity estimates to be equal for the two groups across all symptoms (Model A). For each symptom, another model is fitted that allows the discrimination and severity estimates for the symptom to differ across any two waves, and constrains discrimination and severity estimates of all the remaining symptoms to be equal (Model B). The difference in the log-likelihoods ( $l$ ) of Model A and Model B ( $G^2 = -2(l_{\text{Model A}} - l_{\text{Model B}})$ ) provides an omnibus test ( $df = 2$ ) of whether there is DIF for the discrimination and/or severity estimate for this symptom. If significant, follow-up tests can be conducted to identify whether DIF is present in discrimination or severity estimates.

In conducting DIF analyses that involve tests of multiple group differences across multiple symptoms, it is important to control for the statistical risk of reaching false conclusions. We employed the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), using methods described by Thissen et al. (2002). This procedure adjusts decisions about significance while taking into account the number of significant findings across tests and has been shown to maintain more power than procedures such as the Bonferroni procedure (Benjamini and Hochberg, 1995). A priori, we decided that differences greater than or equal to 0.25 in symptom severity were clinically meaningful. As described by Steinberg and Thissen (2006) a difference of 0.25 can be interpreted as one quarter “standard unit difference between the values of the [underlying] trait necessary to have a 50–50 chance of responding positively in one group compared to another” (p.405–406). This may be considered to be a small effect size (Cohen, 1988). A similar metric is not available for differences in item discrimination parameters. In order to evaluate whether a statistically significant discrimination parameter is also clinically significant we will visually inspect the magnitude of the differences between plotted item response functions (Steinberg and Thissen, 2006).

**2.3.4. Analytic samples**—Although data for mFTQ and DSM-IV were available for all five waves, the NDSS was available only during the final 3 waves of assessment. Thus, the analytic sample was restricted to adolescents who reported at least some cigarette use in the last 30 days during W3, W4, and W5 of assessment:  $n = 253, 241, \text{ and } 296$  respectively.

## 3. Results

### 3.1. Descriptive analyses

A detailed description of the entire sample has been reported (Kandel et al., 2007). We provide a summary of the analytical sub-samples of smokers reporting smoking at least one cigarette in the last 30 days at each wave who were included in the current analysis. At W3, adolescents who smoked were on average 15.9 years old ( $S.D. = 1.3$ ), 50% were female and the racial ethnic distribution was 33% non-Hispanic white, 30% non-Hispanic African American, 37% Hispanic. At W4, adolescents who smoked were on average 16.4 years old ( $SD = 1.2$ ), 51% were female and the racial ethnic distribution was 32% non-Hispanic white, 29% non-Hispanic African American, and 39% Hispanic. At W5, adolescents who smoked were on average 16.9

years old ( $SD = 1.2$ ), 46% were female and the racial ethnic distribution was 33% non-Hispanic white, 30% non-Hispanic African American, 37% were Hispanic. By W5, the distribution of number of cigarettes smoked lifetime was 1 cigarette or less 5.4%, 2–5 cigarettes 10.8%, 6–15 cigarettes 10.8%, 16–25 cigarettes 13.9%, 25–99 cigarettes 20.6% and 100 or more cigarettes 38.5%. The frequency of adolescents reporting smoking on all previous 30 days for W5 was 21%.

### 3.2. DSM-IV symptoms

**3.2.1. Unidimensionality and local dependence**—EFA of tetrachoric correlations using robust weighted least squares analysis suggested a single strong primary factor underlying ND criteria across each of the three waves. Given the DSM-IV criteria are based on symptom reports, we will refer to criteria as ‘symptoms’ hereafter to facilitate discussion that can continue to refer to individual symptoms across all instruments. Eigenvalues for the first factor ranged from 3.3–4.2 and eigenvalues for the second factor ranged from 0.8–1.3. The average ratio of first to second factors was 3.7 ( $SD = 1.03$ ), and all symptoms loaded  $> 0.44$  on the primary dimension. There was no symptom pair loading significantly (e.g.  $> 0.44$ ) on secondary factors to suggest local dependencies. These results suggest a strong primary dimension underlying responses to the DSM-IV symptoms. The primary dimension accounted for 37.3%, 38.3%, and 38.1% of the common variance for W3–W5 respectively.

### 3.3. Modified Fagerstrom (mFTQ)

**3.3.1. Unidimensionality and local dependence**—EFA of tetrachoric correlations using robust weighted least squares analysis suggested a single strong primary factor underlying mFTQ criteria across each of the three waves. Eigenvalues for the first factor ranged from 3.1–3.4, and eigenvalues for the second factor ranged from 1.2–1.6. The average ratio of first to second factors was 2.33 ( $SD = 0.37$ ), and all symptoms had average loadings  $> 0.32$  on the primary dimension. The primary dimension accounted for 40.4%, 46.5%, and 45.4% of the common variance for W3–W5 respectively.

### 3.4. Nicotine Dependence Syndrome Scale (NDSS)

**3.4.1. Unidimensionality and local dependence**—EFA of polychoric correlations using robust weighted least squares analysis suggested a single strong primary factor underlying NDSS criteria across each of the three waves. Eigenvalues for the first factor ranged from 9.7–10.6, and eigenvalues for the second factor ranged from 1.1–1.5. The average ratio of first to second factors was 7.59 ( $SD = 1.61$ ), and all symptoms had average loadings  $> 0.50$  on the primary dimension. The correlations of the first two factors were 0.74, 0.63, and 0.69 for W3–W5, respectively. The primary dimension accounted for 47.9%, 38.1%, and 54% of the common variance for W3–W5, respectively. However, in estimating additional factors, three symptoms [symptoms 3, 10, 16] were estimated to have loadings exceeding 1.0 and/or generated negative residual variance suggesting problems with fitting more than one factor to these data. Upon inspection, each of these symptoms mirrored content and had high shared correlations with adjacent symptoms (e.g.  $r$ 's = 0.79, 0.92, 0.92 for symptoms 2, 9, and 17 respectively) and was thus dropped from further analyses to minimize potential local dependence.

**3.4.2. Effectiveness of NDSS response options**—Prior to examining fit to an item response model, the 5-level response format of the NDSS was evaluated. We examined how the probability of endorsing each of the 5-level response options for each NDSS symptom changed with increasing ND using a nonparametric model (Ramsay, 2000) that, unlike parametric models, makes few assumptions about the relationship between levels of dependence and symptom responses. Examination of the curves suggested good performance



for the initial NDSS options (either ‘not very true’ or ‘fairly true’), but little further differentiation among higher options (‘very true’ and ‘extremely true’) in these samples. For many symptoms, the first option (‘not true at all’) was the most likely choice for most respondents and the multiple additional response options did not produce additional reliable separation of respondents within higher levels of ND in this population. These results suggested: a) accurate modeling of all five response options in these data was not likely to yield replicable results, and b) dichotomizing NDSS responses would not result in a significant loss of information or decrease reliability of the rank-order of adolescents ND in these samples. A reanalysis of the primary dimension of the 14 dichotomized NDSS symptoms again suggested a strong single factor accounting for 60.7%, 51.6%, and 51.5% of the common variance in W3–W5 respectively. Therefore, we proceeded with NDSS symptoms using a dichotomized format (i.e. endorsement of ‘not true at all’ was coded negatively and endorsement other options was coded positively).

### 3.5. Combined Index of Nicotine Dependence Symptoms

**3.5.1. Unidimensionality and local dependence**—CFA of polychoric correlations using robust weighted least squares analysis suggested a single strong primary factor underlying all 28 symptoms (7 DSM-IV; 7 mFTQ; 14 NDSS) across each of the three waves. Eigenvalues for the first factor ranged from 10.6–12.0, and eigenvalues for the second factor ranged from 1.1–2.7. The average ratio of first to second factors was 4.33 ( $SD = 0.38$ ), and all symptoms had average loadings  $> 0.50$  on the primary dimension in both unidimensional and hierarchical models that controlled for variability due to secondary factors. Table 1 presents fit indices for each of the three examined models. Across waves, the hierarchical model provided the strongest fit indices suggesting that the three measures combine to index a primary dimension. The multidimensional model with three related factors suggested strong correlations among factors that were on average 0.81 ( $S.D. = 0.11$ ), 0.71 ( $S.D. = 0.04$ ), and 0.74 ( $S.D. = 0.12$ ) for waves 3–5 respectively. Although a model that includes multiple dimensions fits better than a unidimensional model, results from the CFA suggest that a strong primary dimension predominates and that the primary relationship between symptoms and levels of ND are not significantly affected by multidimensionality in these data. The loading of each symptom on the primary dimension of ND from unidimensional models changed on average only by  $-0.03$  ( $S.D. = 0.05$ ) after controlling for the influence of multiple secondary dimensions in the hierarchical CFA model.

**3.5.2. Choosing a parametric item response model**—Given evidence for the strength of a primary dimension of ND underlying responses to the combined set of 28 items, we proceeded with a unidimensional item response model. We compared the fit of two IRM models: (a) a model that required all symptoms to discriminate similarly; and (b) a model that allowed each symptom to have different levels of discrimination. Evaluation of the log likelihoods suggested that allowing the discrimination parameter to vary in the two-parameter model improved the fit to the data ( $G^2 = 153.16, 173.38, 146.56$  at waves 3–5;  $df = 27, p's < 0.001$ ).

**3.5.3. Fit of the symptoms to the model**—Table 2 lists the severity and discrimination parameter estimates for each of the 28 ND symptoms. Positive values reflect symptoms likely to be endorsed among those with above average levels of ND, and symptoms with negative values reflect symptoms likely to be endorsed among those with below-average levels of ND. To evaluate the fit of each symptom we computed a  $X^2$  fit statistic that evaluated the difference between the observed and model-expected proportion of responses for each symptom across a series of 10 successive intervals of ND (Reise, 1990). Although the statistic typically is distributed as chi-square, we used a Monte Carlo procedure to ensure that we approximated the actual distribution of the fit statistic (Yen, 1981). Overall, the data fit the IRM model well

(see Table 2). All symptoms from all waves had Chi Square fit values within acceptable range except for the ‘Inhale’ symptom from the mFTQ, which had significant misfit at W4 ( $X^2=2.504$ ,  $p < 0.05$ ). However, this symptom showed acceptable fit at W3 and W5 (see Table 2). Similarly NDSS symptoms 1, 2, 12 and 17 indicated misfit at one wave and not another. Therefore, overall there was good fit to the IRM model.

**3.5.4. Replication of the model over time**—The combined index of DSM-IV, mFTQ and NDSS symptoms was evaluated for differences in discrimination and severity estimates over assessments. The estimates for each symptom were examined and tested using a Likelihood Ratio Tests (LRT or  $G^2$ ) of models that systematically constrained all unexamined estimates to be the same. This process allowed each severity estimate and each discrimination estimate to be isolated and tested across waves while controlling for all other symptoms (Thissen, 1991).

When comparing W3 and W4, three symptoms (DSM-IV 3: ‘Impaired Control’, mFTQ 2: ‘Inhale’, and NDSS 15: ‘Amount you smoke has increased’) were identified as having potential differences across waves with a  $G^2$  greater than our critical value of 3.85 (e.g.  $p < 0.05$ ) and were removed. These three items were then reevaluated using the remaining 25 symptoms as an anchor for levels of ND. Two of the three symptoms (mFTQ 2, and NDSS 15) continued to evidence significant DIF with  $G^2 > 3.85$  (See Figure 2). Both symptoms had a slope estimate that differed across W3 and W4 (mFTQ symptom 2:  $a_{w3} = 0.34$ ,  $a_{w4} = 1.92$ ,  $G^2 = 5.4$ ,  $df = 1$ ,  $p < 0.03$ ). Similarly, NDSS symptom 15 had a slope estimate that differed across W3 and W4 ( $a_{w3} = 1.83$ ,  $a_{w4} = 3.04$ ,  $G^2 = 4.7$ ,  $df = 1$ ,  $p < 0.04$ ). Figure 1 displays the item response functions for symptoms with statistically significant slope differences. From this figure it is apparent that the mFTQ ‘Inhale’ item had a greater than 50% probability of being observed across a full range of ND scores and thus discriminated poorly across both waves. The NDSS symptom 15 discriminated effectively across waves and appeared to be slightly less severe in W4 relative to W3. However, within ranges where this symptom is likely to be observed ( $ND > -0.04$ ), any differences in the assessment of levels of ND across these waves would be small (Cohen’s  $d$  ranged 0.01–0.13) and not deemed clinically significant.

In comparing W4 and W5, only one item, NDSS symptom 9 (‘Smoke same on weekends as on weekdays’), evidenced significant DIF. It was removed and tested using the remaining 27 symptoms as an anchor. This item was estimated as being more severe at W4 than at W5 ( $b_{w4} = -0.34$ ,  $b_{w5} = -0.89$ ,  $d = 0.55$ ,  $G^2 = 11.5$ ,  $df = 1$ ,  $p < 0.001$ ).

In comparing W3 and W5, three NDSS (1, 5, and 14) symptoms evidenced significant DIF, were removed, and tested using the remaining 25 symptoms as an anchor. NDSS 1 (‘Need to smoke to feel less restless’) was estimated to be more discriminating at W3 than at W5 ( $a_{w3} = 2.57$ ,  $a_{w4} = 1.52$ ,  $G^2 = 6.2$ ,  $df = 1$ ,  $p < 0.02$ ). Figure 1 displays the effect of the increased discrimination for NDSS 1 in W3 relative to W5. The amount of difference in discrimination caused the symptom to be slightly more severe at W5 in the upper ranges of ND where this symptom is likely to be observed and did not result in clinically significant levels of DIF within this range of ND. NDSS symptom 5 (‘Avoid places that do not allow smoking’) was estimated as a less severe symptom at W3 than at W5 ( $b_{w3} = 0.44$ ,  $b_{w5} = 0.81$ ,  $d = .37$ ,  $G^2 = 4.6$ ,  $df = 1$ ,  $p < 0.04$ ). NDSS symptom 14 (‘Number you smoke not affected much by other things’) was estimated as a more severe symptom at W3 than at W5 ( $b_{w3} = -0.71$ ,  $b_{w5} = -1.40$ ,  $d = .69$ ,  $G^2 = 4$ ,  $df = 1$ ,  $p < 0.05$ ). Overall, item performances were quite stable over time.

To summarize the performance of each measure using the same metric, we computed a test information function to gauge the precision of the DSM-IV, mFTQ, and NDSS as a function of levels of ND. The test information function is a direct reflection of the standard error of each measure and provides a continuous view of where on the latent continuum each measure

performs best. The test information function also reflects the total information from each measure and thus measures with more items (e.g. NDSS) potentially will contribute more information about levels of ND than measures with fewer items (e.g. DSM-IV and mFTQ). Figure 2 displays the test information function for the three measures observed at the final assessment (W5). Although the test information functions were similar across waves, we present results from W5 given the increased variation in the range of levels of nicotine dependence in this assessment wave. Overall the NDSS symptoms were well matched to the range of ND observed by the majority of teens in this sample. However, for teens at the highest levels of ND, the DSM-IV appeared to add unique information. The mFTQ had several items that were not as strongly discriminating as the other measures and also had significant overlap in assessing the higher ranges of ND above traditional DSM-IV diagnostic thresholds.

Using the item parameters from the combined index of ND, we were able to link each set of 7 DSM-IV, 7 mFTQ, and the 14 NDSS items so that raw summed total scores could be compared directly. Summed-score linking (Lord and Wingersky, 1984) involves first regrouping the items for each scale. Then for each scale, an average a posteriori (EAP) IRT score is estimated for each possible summed total score (Dorans, 2007; Orlando et al., 2000; Thissen et al., 1995). We used a program IRTScore (Flora and Thissen, 2002) to estimate the average IRT score (EAP) corresponding to each summed total score for each scale (see Table 3). Summed scores from each measure can be translated by matching the corresponding IRT score across scales. For example, a DSM-IV diagnosis is assigned with a summed score of 3 or more criteria. From Table 3 we find that a DSM-IV score of 3 corresponds to an IRT score of  $\geq 0.58$ . To find the corresponding score on the mFTQ, the nearest IRT score  $\geq 0.58$  is estimated to be a score of  $\geq 4$ . The same IRT score would be  $\geq 11$  on the NDSS.

**3.5.5. Concurrent Validity**—Figure 3 shows the linear relationship (polyserial correlations = .66, .63, .73,  $p$ 's < 0.05) between estimated levels of ND and levels of current smoking across each wave of assessment. At W5, the average level of ND among those smoking less than one, one, two to five, and more than five cigarettes each day was  $-0.91$  ( $S.D. = 0.69$ ),  $-0.43$  ( $S.D. = 0.75$ ),  $0.39$  ( $S.D. = 0.68$ ),  $1.03$  ( $S.D. = 0.60$ ), respectively. The changes in levels of ND suggest a strong relationship between self-reported smoking behavior and increasing levels of ND.

**3.5.6. Patterns of symptoms observed across levels of ND**—We examined patterns of responding to DSM-IV symptoms both above and below diagnostic threshold (e.g.  $\geq 3$  DSM-IV symptoms). There are a number of ways to get each total score on the DSM-IV and each pattern may have different levels of associated ND. To examine the relationship between the DSM-IV and estimated levels of ND we evaluated the distribution of ND scores among adolescents with the same number of DSM-IV symptoms. The distribution of ND scores is represented in box and whisker plots (see Figure 4) in which the boxes reflect the bulk of observed scores (25<sup>th</sup>–75<sup>th</sup> percentiles) that has a solid line representing the median, and whiskers extend to the range of observed scores. Overall, the median estimated a posteriori ND scores increased as the number of observed DSM-IV symptoms increased. In addition, there was considerable variability in levels of ND among individuals with the same number of DSM-IV symptoms, particularly among adolescents below diagnostic threshold.

Although a continuous count of DSM-IV symptoms was related to increases in levels of ND, it appears that considerable heterogeneity exists within each level of observed symptoms and further refinement of the rank-order of adolescents below a diagnostic threshold is possible using symptoms from other measures. Of those teens reporting only one symptom, the most frequent pattern included either Tolerance alone (40%) or Impaired Control (35.5%) alone. Although there are twenty-one possible DSM-IV symptom pairs, we observed only 10 patterns at W5. Of the patterns, the most common (25%) was a combination of Impaired Control and

Tolerance, followed by Tolerance and Withdrawal (19.6%), and then Impaired Control and Difficulty Quitting (17.6%).

Using model estimates we expect adolescents with a DSM-IV diagnosis to primarily endorse symptoms with the lowest severity estimates (e.g. b parameters in Table 2). Although there are 99 possible combinations of symptoms that can lead to a diagnosis of Nicotine Dependence, we observed only 30 unique patterns in these data. The two most common patterns observed among adolescents with three symptoms were distinguished by the endorsement of Impaired Control and Withdrawal along with either Tolerance ( $M_{ND} = 0.20$ ;  $S.D. = 0.31$ ) or Difficulty Quitting ( $M_{ND} = 0.84$ ;  $S.D. = 0.45$ ). These two symptom patterns were associated with significantly different levels of ND ( $t(14) = 3.29$ ,  $p < 0.01$ ). The most common pattern of symptoms leading to a diagnosis included the three least severe symptoms of Impaired Control, Tolerance, and Withdrawal (58.9%).

#### 4. Discussion

This study evaluated the interrelationships of three measures of nicotine dependence in a sample of teens evaluated repeatedly over a two-year period. Each instrument appeared to measure a primary single dimension of nicotine dependence. When combined, the three instruments formed a stable measure of nicotine dependence that had good reliability and broad coverage of the nicotine dependence continuum. The DSM-IV based interview and the mFTQ had significant overlap in assessing the higher ranges of ND above traditional diagnostic thresholds. By contrast, the NDSS provided items that extended the reach of the DSM-IV and mFTQ by assessing levels of ND below diagnostic thresholds, where most adolescents in this study were observed. Despite differences in the overall level of ND targeted by each instrument, there was considerable similarity in the levels of ND associated with items that tapped similar content. By combining items from these three instruments, a level of information was achieved that exceeded any one instrument alone. These complementary assessment instruments may improve understanding of adolescents' progression along a continuum of nicotine dependence severity over time by increasing the reliability of overlapping content and broadening the coverage of levels of ND. In addition, with an increasingly reliable understanding of the patterns of symptoms associated with each level of ND, longitudinal designs can be used to begin to study symptom expression over time, test hypotheses regarding temporal sequencing of symptoms, and evaluate how patterns of increasing or decreasing symptoms relate to nicotine exposure.

The DSM-IV provided an effective anchor for evaluating ND and analyses support the stability of the construct underlying the DSM-IV among teens. The severity of DSM-IV symptoms followed an expected pattern as described in Kandel et al. (2007). The DSM-IV index performed similarly across waves and showed no significant changes in the effectiveness of symptoms across waves. Tolerance, impaired control, and withdrawal symptoms mark a level of ND below the diagnostic threshold, which is delineated most closely by self-reported difficulty quitting. This pattern of symptoms provides additional evidence of difficulty quitting being a hallmark symptom of DSM-IV defined Nicotine Dependence syndrome. Given that the DSM-IV focuses on diagnosis of nicotine dependence, a substance use disorder, rather than indexing a continuum of ND, it was not surprising to find that the DSM-IV instrument provided limited information about ND severity among adolescents with low levels of smoking. For example, 50% of the information provided by the DSM-IV occurs within a range of ND observed among the top 10% of teens in this sample. These results validate the DSM-IV index as a reliable and valid index of levels of ND observed among more teens with significant exposure to nicotine.

Like the DSM-IV, the majority of the mFTQ items target levels of ND that are significantly higher than the level experienced by this sample of adolescents. Of the seven mFTQ items, the 'Inhale' symptom was endorsed by >95% of adolescents, was not consistently endorsed across waves, and provided limited additional information about low levels of ND. The 'Difficulty Refraining' and 'Smoke more in the first two hours' symptoms did not discriminate well among levels of ND. Although endorsement of these symptoms was similar to other mFTQ symptoms, those who endorsed 'Difficulty Refraining' and/or 'Smoke more in the first two hours' symptoms were not necessarily those teens with the highest levels of ND. Thus, 3 of the 7 mFTQ symptoms did not contribute much information about levels of ND in this sample. Across waves, the high severity of mFTQ items placed them within reach of only a small minority of teens and sample limitations may have reduced the ability to establish stable estimates of item characteristics. The mFTQ symptoms have demonstrated good measurement properties in samples of heavier smoking teens, and thus further exploration of the role of the mFTQ symptoms in describing levels of ND is needed. The mFTQ did not appear to add information about ND above and beyond the other measures in this sample of light smokers.

The NDSS provided broad coverage of levels of ND observed in this sample but only a small percentage (12%) of information within the highest levels (e.g. >90<sup>th</sup> percentile) of ND. It is possible that a sample with higher levels of smoking might allow for more stable estimates of additional response options within the NDSS and thereby increase coverage of this range of ND. Adding reliable response options would increase the utility of the NDSS for use among smokers with higher levels of ND. Given our limited ability to evaluate the reliability of all responses to the NDSS, a more quasi-continuous response scale for each item may still make sense for a population with greater dependence. With the dichotomized NDSS, smoking regularly, craving, and concern about running out of cigarettes were among the most discriminating NDSS symptoms. Several of the NDSS symptoms demonstrated some unstable relationships with levels of ND. However, instabilities in the symptoms were not observed across all waves, and the majority of NDSS symptoms were stable over time.

The combined symptoms index a range of ND and provide a means of evaluating the consistency of symptoms from different measures that target similar content. For example, each of the three instruments assesses Tolerance. The DSM-IV inquired about smoking more without feeling nauseated or smoking more to feel satisfied (DSM-IV 1 = 0.34). The mFTQ inquires about the quantity of smoking (mFTQ 1 = -0.30), and the NDSS inquires about increasing the amount smoked over time (NDSS 15 = -0.32) or being able to smoke more (NDSS 17 = -0.22). Each of these approaches to the construct of Tolerance measured a slightly different level of severity and thereby increases the range within which symptoms of tolerance can be scaled. Similarly, withdrawal avoidance can be manifested by either smoking to avoid 'feeling bad' (DSM-IV 2 = 0.70), smoking within 30 minutes of awakening (mFTQ 3 = 1.45), or smoking to feel less restless (NDSS 1 = 0.05). These different operationalizations provide a broad range (0.05–1.45) of levels of ND to both increase the reliability of rank-ordering adolescents and provide a benefit of multiple items to broaden assessment of these constructs.

The placement of items from three different measures of ND allows for the mapping of a typical pattern of symptoms within levels of nicotine dependence severity. Of the available self-descriptions, the least severe symptom involves a willingness to inhale cigarette smoke (mFTQ 2). This exposure to nicotine self-administration may set a sufficient course for adolescents who at the next level of dependence report unstable patterns of use (NDSS 9) that are influenced by environmental circumstances (NDSS 13). With increased use (mFTQ 1, NDSS 15), adolescents begin to report increased craving to smoke in-between cigarettes (NDSS 2) and report attempts to avoid the discomfort of withdrawal symptoms (NDSS 1). This behavior may be linked to progression to levels of ND that include a loss of control over their use (DSM-IV 3), the experience of difficulty quitting (DSM-IV 4), and the experience of frustration when



environmental circumstances suggest refraining from smoking (NDSS 7, NDSS 5). At the most severe end of the ND continuum, adolescents will be highly motivated to maintain their smoking behavior and are unlikely to be deterred even, for example, if a physical illness (mFTQ 6) keeps them in bed all day, and are willing to disengage from important activities to accommodate smoking (DSM-IV 6). This arrangement of observed symptoms provides a hint at possible course for movement across levels of nicotine dependence. The ND continuum was replicated across the three waves of assessments despite variability in smoking behaviors among teens over time. These results do provide evidence for an organizing dimension of nicotine dependence that here is used to clarify interrelationships among heterogeneous symptoms within specific levels of ND.

The identification of an organization of symptoms along a common continuum of ND facilitates research efforts to understand the construct of ND and manifestation of ND symptoms across a range of tobacco use in adolescents. Evidence for a continuum supports the utility of moving beyond single diagnostic thresholds when examining ND in adolescents. By using a model based on item response theory, we provide an efficient description of the continuum that facilitates research efforts to select sets of items that effectively capture desired contents and efficiently scale levels of nicotine dependence among adolescents. The process of selecting these sets of symptoms may change with the goal of the measurement. For example, genetic research or clinical outcomes researchers may wish to capture maximum amounts of heterogeneity to isolate patterns or acute changes in symptoms in a sample and thus may wish to select a broad range of dependence symptoms. Conversely, an epidemiologist may wish to refine measures by selecting reduced sets of symptoms to maximize efficient classification of those with and without clinically significant levels of nicotine dependence. In either case these initial analyses provide preliminary guidance with regard to selecting from available symptoms. Given that we used the same scale for the items and individuals, tabled values of severity (b parameters) can be used to identify symptoms that are likely to distinguish among adolescents at various levels of ND. For example, figure 4 shows the bulk of ND scores (25<sup>th</sup>–75<sup>th</sup> percentile) for the adolescents with 3+ DSM-IV symptoms ranges from 0.25–1.75. Thus, items with severity estimates just below and within this range are likely to separate adolescents above and below this level of ND. For example, mFTQ items ‘First is most difficult to give up’ (W5 b = 0.98), ‘Time to first cigarette’ (W5 b = 1.45) and an NDSS item relating ‘Avoiding places that do not allow smoking’ (W5 b = 0.81) all begin to be observed within this range of ND. The DSM-IV items closest to this range of ND are ‘Unable to quit’ (W5 b = 0.87) and ‘Use despite physical or psychological problems’ (DSM-IV 7: W5 b = 1.52). The symptoms from the DSM-IV, NDSS, and mFTQ are not a finite set and more work needs to be done to increase available symptoms to characterize these important regions of the continuum and to validate the generalizability of the current symptom patterns. By using a model-based methodology, we provide a practical method for linking multiple existing measures of ND and provide a basis for beginning to develop larger sets or banks of items to be used across a broad range of research applications. It will also be important to implement this type of analysis on smokers of different ages at different levels of dependence.

Although this set of symptoms provided significant information about the rank-order of the level of ND in this sample of adolescents, it is uncertain whether the same level of measurement precision is possible among teens with higher levels of ND severity. Enhancing the phenotypic descriptions of the upper regions of ND is needed. We were able to demonstrate significant variability in levels of ND among adolescents with the same number of DSM-IV symptoms. Although we observed variability in levels of ND associated with particular patterns of responses leading to the same DSM-IV scores, more research is needed to evaluate the utility of further elaborating the characteristics of adolescents with levels of ND below diagnostic threshold. Recently completed analyses are identified potential developmental sequences in symptoms of dependence (Kandel et al., submitted for publication).

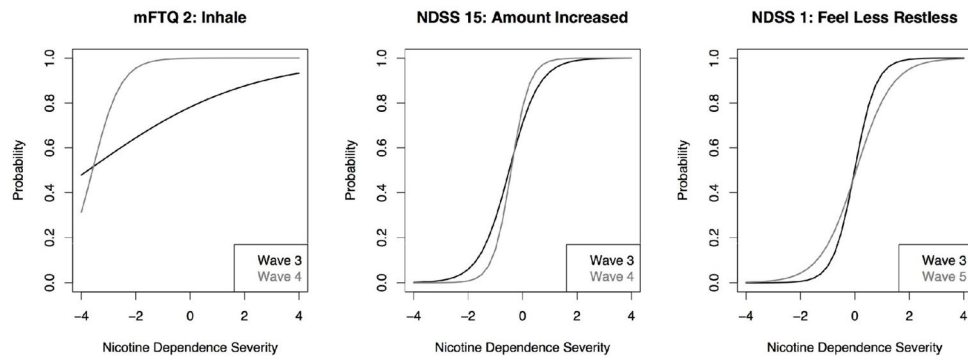
This study has several limitations including the adoption of a single organizing construct for modeling individual symptoms. ND is a complex construct defined by multiple abstract criteria and alternative models may emerge to capture this complexity. Despite the complex and multidimensional nature of ND, researchers and clinicians typically use a single severity score for making predictions or clinical decisions (e.g. benefits of nicotine replacement). As with all complex constructs, there is strong potential for multidimensional measures to further inform phenotypic research. As our research base grows to better describe smoking progression, novel approaches to evaluating symptom domains may be added to existing facets of ND. An additional limitation of this study was the relatively small number of smokers who could be evaluated across all waves of assessment. This prevented a thorough examination of longitudinal sequencing of all symptom trajectories. Longitudinal studies are needed to better track the trajectories of symptom development with fluctuations in smoking rates and to document transitions along a continuum of ND. Future examination of this important data could examine the roles that gender, socioeconomic, racial/ethnic and other individual differences play in determining symptom endorsements. Refining the concept of nicotine dependence in adolescence is crucial to research, prevention and therapeutic efforts. An improved measure of the phenotype is needed for epidemiological and genetic studies so as to better identify risk factors and consequences of nicotine dependence. Better measures of the phenotype are also needed to determine, provide and evaluate preventive and therapeutic efforts.

## References

- Alterman AI, Gariti P, Cook TG, Cnaan A. Nicodermal patch adherence and its correlates. *Drug Alcohol Depend* 1999;53:159–165. [PubMed: 10080041]
- Association, AP. *Diagnostic and statistical manual of mental disorders* (4th ed. Washington, DC: American Psychiatric Association; 1994.
- Benjamini Y, Hochberg Y. Controlling false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 1995;(57):289–300.
- Bentler PM. Comparative fit indices in structural models. *Psychological Bulletin* 1990;107(2):238–246. [PubMed: 2320703]
- Bentler PM, Bonnett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin* 1980;88:588–606.
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In: Novick, FMLMR., editor. *Statistical theories on mental test scores*. Reading: Addison-Wesley; 1968.
- Breslau N, Johnson EO. Predicting smoking cessation and major depression in nicotine-dependent smokers. *Am J Public Health* 2000;90(7):1122–1127. [PubMed: 10897192]
- Browne; Cudeck. *Testing structural equation models*. Newbury Park: Sage; 1993.
- Campbell IA, Prescott RJ, Tjeder-Burton SM. Transdermal nicotine plus support in patients attending hospital with smoking-related diseases: A placebo-controlled study. *Respir Med* 1996;90(1):47–51. [PubMed: 8857326]
- Clark DB, Wood DS, Martin CS, Cornelius JR, Lynch KG, Shiffman S. Multidimensional assessment of nicotine dependence in adolescents. *Drug Alcohol Depend* 2005;77(3):235–242. [PubMed: 15734223]
- Cohen, J. *Statistical power for the behavioral sciences* (second edition). Vol. 2. Hillsdale, N.J.: Lawrence-Erlbaum Associates, Inc; 1988.
- Cohen LM, Myers MG, Kelly JF. Assessment of nicotine dependence among substance abusing adolescent smokers: A comparison of the dsm-iv criteria and the modified fagerstrom tolerance questionnaire. *Journal of Psychopathology and Behavioral Assessment* 2002;24(4):225–233.
- Crowley TJ. Adolescents and substance-related disorders: Research agenda to guide decisions on diagnostic and statistical manual of mental disorders, fifth edition (DSM-V). *Addiction* 2006;101 (Suppl 1):115–124. [PubMed: 16930168]

- Crowley TJ, Mikulich SK, Ehlers KM, Whitmore EA, MacDonald MJ. Validity of structured clinical evaluations in adolescents with conduct and substance problems. *J Am Acad Child Adolesc Psychiatry* 2001;40(3):265–273. [PubMed: 11288767]
- DiFranza JR. Implications of the autonomy theory of nicotine dependence. *MedGenMed* 2002;4(3):8. [PubMed: 12466751]
- DiFranza JR, Rigotti NA, McNeill AD, Ockene JK, Savageau JA, St Cyr D, Coleman M. Initial symptoms of nicotine dependence in adolescents. *Tob Control* 2000;9(3):313–319. [PubMed: 10982576]
- DiFranza JR, Wellman RJ. A sensitization-homeostasis model of nicotine craving, withdrawal, and tolerance: Integrating the clinical and basic science literature. *Nicotine Tob Res* 2005;7(1):9–26. [PubMed: 15804674]
- Dorans NJ. Linking scores from multiple health outcome instruments. *Qual Life Res* 2007;16(Suppl 1): 85–94. [PubMed: 17286198]
- Edwards G, Gross MM. Alcohol dependence: Provisional description of a clinical syndrome. *Br Med J* 1976;1(6017):1058–1061. [PubMed: 773501]
- Etter JF, Duc TV, Perneger TV. Validity of the fagerstrom test for nicotine dependence and of the heaviness of smoking index among relatively light smokers. *Addiction* 1999;94(2):269–281. [PubMed: 10396794]
- Fagerstrom KO. Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addictive Behaviors* 1978;3:235–241. [PubMed: 735910]
- Flora, D.; Thissen, D. Irtscore: Item response theory score approximation software. Chapel Hill, NC: L.L. Thurstone Psychometric Laboratory; 2002.
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The fagerstrom test for nicotine dependence: A revision of the fagerstrom tolerance questionnaire. *British Journal of Addiction* 1991;86:1119–1127. [PubMed: 1932883]
- Kandel DB, Hu MC, Griesler PC, Schaffran C. On the development of nicotine dependence in adolescence. *Drug Alcohol Depend* 2007;91(1):26–39. [PubMed: 17553635]
- Kandel DB, Schaffran C, Griesler P, Samuolis J, Davies M, Galanti R. On the measurement of nicotine dependence in adolescence: Comparisons of the mFTQ and a DSM-IV-based scale. *J Pediatr Psychol* 2005;30(4):319–332. [PubMed: 15863429]
- Lord, F. Applications of item response theory to practical testing problems. Hillsdale, NJ: Erlbaum; 1980.
- Lord FM, Wingersky MS. Comparison of true score and equipercentile observed-score ‘equatings’. *Applied Psychological Measurement* 1984;8:453–461.
- Mikulich SK, Hall SK, Whitmore EA, Crowley TJ. Concordance between dsm-iii-r and dsm-iv diagnoses of substance use disorders in adolescents. *Drug Alcohol Depend* 2001;61(3):237–248. [PubMed: 11164688]
- Orlando M, Sherbourne CD, Thissen D. Summed-score linking using item response theory: Application to depression measurement. *Psychol Assess* 2000;12(3):354–359. [PubMed: 11021160]
- Patten CA, Ames SC, Ebbert JO, Wolter TD, Hurt RD, Gauvin TR. Tobacco use outcomes of adolescents treated clinically for nicotine dependence. *Arch Pediatr Adolesc Med* 2001;155(7):831–837. [PubMed: 11434853]
- Piper ME, Piasecki TM, Federman EB, Bolt DM, Smith SS, Fiore MC, Baker TB. A multiple motives approach to tobacco dependence: The wisconsin inventory of smoking dependence motives (WISDM-68). *J Consult Clin Psychol* 2004;72(2):139–154. [PubMed: 15065950]
- Prokhorov AV, Koehly LM, Pallonen UE, Hudmon KS. Adolescent nicotine dependence measured by the modified fagerstrom tolerance questionnaire at two time points. *Journal of Child and Adolescent Substance Abuse* 1998;7(4):35–47.
- Ramsay, JO. Testgraf: A program for the graphical analysis of multiple choice test and questionnaire data. 2000. Unpublished manual
- Reckase MD. The past and future of multidimensional item response theory. *Applied Psychological Measurement* 1997;21:25–36.
- Reise SP. A comparison of item- and person-fit methods of assessing model-data fit in irt. *Applied Psychological Measurement* 1990;14:127–137.

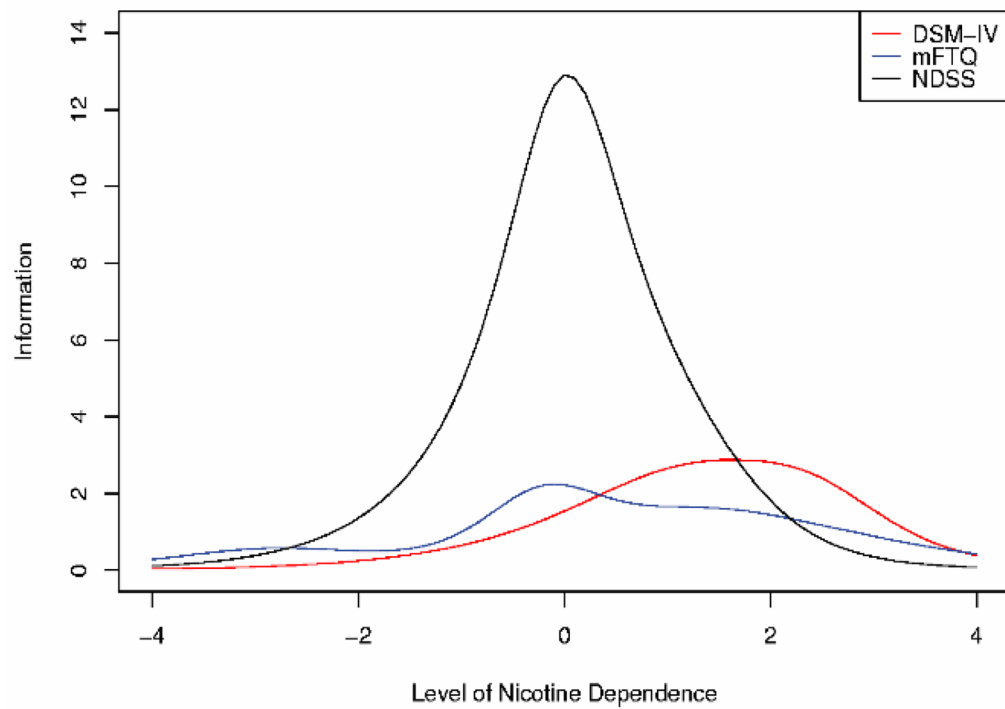
- Reise SP, Morizot J, Hays RD. The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Qual Life Res* 2007;16(Suppl 1):19–31. [PubMed: 17479357]
- Rizopoulos D. Ltm: An r package for latent variable modeling and item response analysis. *Journal of Statistical Software* 2006;17(5):1–10.
- Rojas NL, Killen JD, Haydel KF, Robinson TN. Nicotine dependence among adolescent smokers. *Arch Pediatr Adolesc Med* 1998;152(2):151–156. [PubMed: 9491041]
- Shiffman S, Waters A, Hickcox M. The nicotine dependence syndrome scale: A multidimensional measure of nicotine dependence. *Nicotine Tob Res* 2004;6(2):327–348. [PubMed: 15203807]
- Sledjeski EM, Dierker LC, Costello D, Shiffman S, Donny E, Flay BR. Predictive validity of four nicotine dependence measures in a college sample. *Drug Alcohol Depend* 2007;87(1):10–19. [PubMed: 16930859]
- Steiger JH. Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research* 1990;25:173–180.
- Steinberg L, Thissen D. Using effect sizes for research reporting: Examples using item response theory to analyze differential item functioning. *Psychol Methods* 2006;11(4):402–415. [PubMed: 17154754]
- Strong DR, Brown RA, Ramsey SE, Myers MG. Nicotine dependence measures among adolescents with psychiatric disorders: Evaluating symptom expression as a function of dependence severity. *Nicotine Tob Res* 2003a;5(5):735–746. [PubMed: 14577990]
- Strong DR, Kahler CW, Abrantes AM, MacPherson L, Myers MG, Ramsey SE, Brown RA. Nicotine dependence symptoms among adolescents with psychiatric disorders: Using a rasch model to evaluate symptom expression across time. *Nicotine Tob Res* 2007;9(5):557–569. [PubMed: 17454712]
- Strong DR, Kahler CW, Ramsey SE, Brown RA. Finding order in the dsm-iv nicotine dependence syndrome: A Rasch analysis. *Drug Alcohol Depend* 2003b;72(2):151–162. [PubMed: 14636970]
- Thissen, D. *Multilog users guide- version 6*. Chicago, IL: Scientific Software, Inc; 1991.
- Thissen, D. *Irtldif v.20b*: Software for the computation of the statistics involved in item response theory likelihood ratio tests for differential item functioning. University of North Carolina; Chapel Hill, NC: 2001. Unpublished manuscript
- Thissen D, Pommerich M, Billeaud K, Williams V. Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement* 1995;19:39–49.
- Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics* 2002;27(1):77–83.
- Wellman RJ, Savageau JA, Godiwala S, Savageau N, Friedman K, Hazelton J, Difranza JR. A comparison of the hooked on nicotine checklist and the fagerstrom test for nicotine dependence in adult smokers. *Nicotine Tob Res* 2006;8(4):575–580. [PubMed: 16920655]
- Westman EC, Behm FM, Simel DL, Rose JE. Smoking behavior on the first day of a quit attempt predicts long-term abstinence. *Arch Intern Med* 1997;157(3):335–340. [PubMed: 9040302]
- Wheeler KC, Fletcher KE, Wellman RJ, Difranza JR. Screening adolescents for nicotine dependence: The hooked on nicotine checklist. *J Adolesc Health* 2004;35(3):225–230. [PubMed: 15313504]
- Yen W. Using simulation results to choose a latent trait model. *Applied Psychological Measurement* 1981;5(5):245–262.
- Young SE, Corley RP, Stallings MC, Rhee SH, Crowley TJ, Hewitt JK. Substance use, abuse and dependence in adolescence: Prevalence, symptom profiles and correlates. *Drug Alcohol Depend* 2002;68(3):309–322. [PubMed: 12393225]
- Young SE, Mikulich SK, Goodwin MB, Hardy J, Martin CL, Zoccolillo MS, Crowley TJ. Treated delinquent boys' substance use: Onset, pattern, relationship to conduct and mood disorders. *Drug Alcohol Depend* 1995;37(2):149–162. [PubMed: 7758404]
- Yu, CY. Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. University of California; Los Angeles: 2002.



**Figure 1.** Item response functions for the three symptoms with statistically significant differences in item discrimination across waves of assessment.

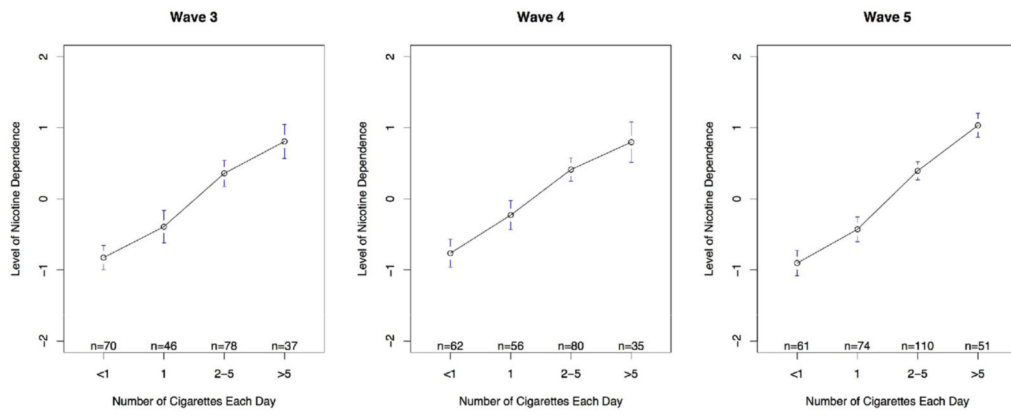


## Test Information Functions for DSM-IV, mFTQ, and NDSS

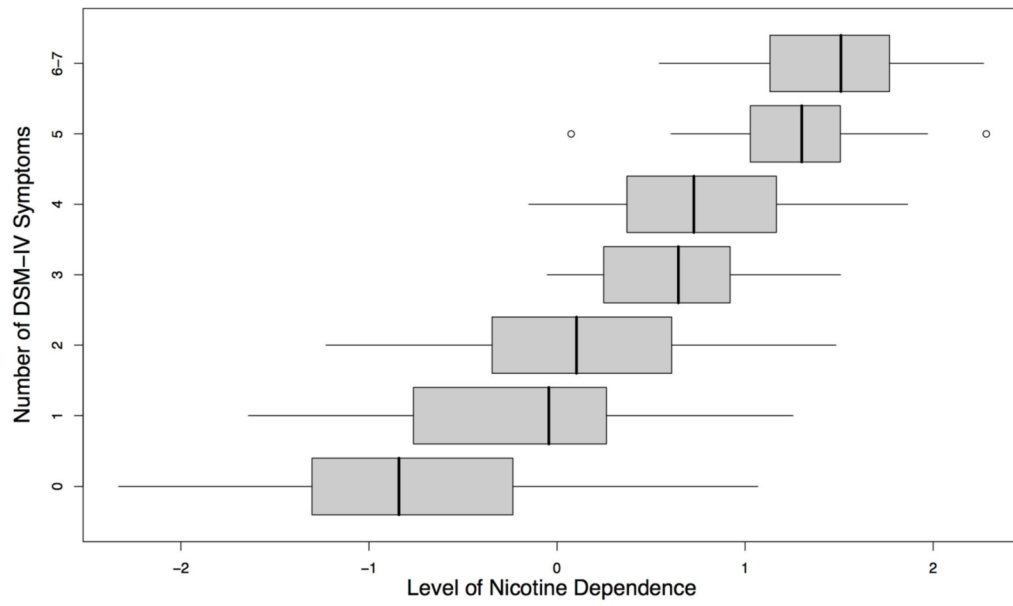


**Figure 2.**

Test information functions for the DSM-IV, mFTQ, and NDSS measures of nicotine dependence at the final wave of assessment (W5). The latent level of nicotine dependence is standardized for this sample to have a mean of zero and a one-unit standard deviation. The amount of information in each instrument is a function of the number and quality of the items.



**Figure 3.** Mean level of nicotine dependence severity for adolescents as a function of the amount of smoking in the last 30 days prior to waves 3–5.



**Figure 4.** Boxplots of estimated a posteriori (EAP) scores for the level of nicotine dependence for each number of DSM-IV symptoms. Solid vertical lines within boxes represent group medians.

**Table 1**

Results from confirmatory factor analysis of 7 DSM-IV criteria, 7 mFTQ items, and 14 NDSS <sup>a</sup> items to assess an unidimensional, a multidimensional (three factors), and a hierarchical bifactor model in which a general factor is evaluated along with three orthogonal factors representing DSM-IV, mFTQ, and NDSS item groups.

	<b>Model A</b>	<b>Model B</b>	<b>Model C</b>
<b>Wave 3</b>	<b>Unidimensional</b>	<b>Multidimensional</b>	<b>Hierarchical</b>
CFI	0.935	0.952	0.967
TLI	0.951	0.965	0.975
RMSEA	0.073	0.062	0.052
<b>Wave 4</b>			
CFI	0.902	0.926	0.948
TLI	0.925	0.943	0.959
RMSEA	0.082	0.071	0.060
<b>Wave 5</b>			
CFI	0.936	0.957	0.969
TLI	0.950	0.967	0.975
RMSEA	0.074	0.061	0.052

<sup>a</sup>Three NDSS items were excluded from the analysis. CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA= Root Mean Square Error of Approximation.

**Table 2**  
Frequency and item response model estimates for each symptom from the Diagnostic and Statistics Manual (DSM-IV), Modified Fagerstrom Tolerance Questionnaire (mFTQ), and Nicotine Dependence Symptom Scale (NDSS).

Symptoms	Wave 3 (n = 253)				Wave 4 (n=241)				Wave 5 (n=296)			
	%	a	b	X <sup>2</sup> Fit	%	a	b	X <sup>2</sup> Fit	%	a	b	X <sup>2</sup> Fit
<i>DSM-IV</i>												
1. Tolerance	45.9	1.20	0.14	0.29	41.4	1.10	0.42	0.39	42.9	1.04	0.34	0.19
2. Withdrawal	35.2	0.79	0.73	0.30	32.0	0.91	1.00	0.37	33.7	1.29	0.70	0.12
3. Impaired control	52.4	1.45	-0.18	0.27	48.3	0.85	0.12	1.60	49.0	0.95	0.05	0.38
4. Unable to quit	24.8	1.54	0.96	0.36	22.8	1.67	1.06	0.05	27.6	1.68	0.87	0.09
5. Great deal of time spent using tobacco	16.0	1.19	1.71	0.09	14.2	1.15	1.90	0.52	15.3	1.16	1.82	0.05
6. Neglect important activities	4.8	0.93	3.64	0.41	4.3	0.76	4.39	0.38	3.4	2.11	2.37	0.43
7. Use despite physical or psychological problems	13.0	1.24	1.89	0.32	19.4	1.20	1.48	0.10	14.3	1.74	1.52	0.09
<i>mFTQ</i>												
1. Daily quantity smoked	55.6	1.96	-0.32	0.22	59.1	2.28	-0.22	0.07	59.9	2.61	-0.30	0.07
2. Inhaling smoke	94.8	0.34 <sup>a</sup>	-3.66	0.43	97.0	1.92 <sup>d</sup>	-2.37	2.50*	97.3	1.50	-2.88	0.14
3. Time to first cigarette	16.0	1.74	1.35	0.21	18.3	1.75	1.27	0.07	17.1	1.53	1.45	0.15
4. First is most difficult to give up	20.5	1.49	1.16	0.13	20.3	1.05	1.56	0.33	26.2	1.50	0.98	0.15
5. Difficulty refraining	21.2	0.29	4.3	0.23	19.8	0.35	4.05	0.22	17.3	0.75	2.30	0.21
6. Smoking if ill in bed	4.8	1.09	3.13	0.10	6.0	1.13	2.83	0.37	6.1	1.31	2.59	0.04
7. Smoking more during the first two hours	6.4	0.65	4.33	2.08	3.9	0.09	37.34	0.32	6.2	0.40	6.98	0.58
<i>NDSS</i>												
1. Need to smoke to feel less restless	47.6	2.57 <sup>d</sup>	-0.02	0.07	53.9	1.87	-0.11	0.43*	49.0	1.52 a	0.05	0.12
2. After a few hours, crave cigarettes	45.0	3.35	0.05	0.06	49.6	2.42	0.06	0.61*	48.0	3.02	0.07	0.08
3. Cravings like in the grip of an unknown force	38.1				36.9				42.6			
4. Feel a sense of control over smoking (reversed)	66.2	1.17	-0.86	0.45	72.0	1.00	-1.09	0.47	64.3	1.29	-0.60	0.33



Symptoms	Wave 3 (n = 253)				Wave 4 (n=241)				Wave 5 (n=296)			
	%	a	b	X <sup>2</sup> Fit	%	a	b	X <sup>2</sup> Fit	%	a	b	X <sup>2</sup> Fit
5. Avoid places that do not allow smoking	39.4	0.88	0.44 <sup>a</sup>	0.25	36.6	0.78	0.79	0.33	33.3	1.18	0.81 <sup>d</sup>	0.10
6. Rather not travel by airplane due to nonsmoking	14.7	2.02	1.34	0.08	16.4	1.66	1.42	0.14	14.6	1.62	1.55	0.26
7. Worry that you will run out of cigarettes	22.5	2.24	0.88	0.11	27.6	2.99	0.71	0.15	27.6	2.92	0.72	0.14
8. Smoke regularly throughout the day	51.3	3.29	-0.18	0.05	57.3	2.86	-0.18	0.07	54.8	2.99	-0.13	0.03
9. Smoke same on weekends as on weekdays	59.6	1.43	-0.52	0.26	58.8	1.48	-0.34 <sup>d</sup>	0.21	68.6	1.46	-0.89 <sup>d</sup>	0.31
10. Smoke same number day to day	59.6				59.1				69.0			
11. Number changes day to day (reversed)	59.6	2.00	-0.38	0.15	59.1	1.93	-0.26	0.07	60.2	1.66	-0.38	0.13
12. Normal not to have another for hours. (reversed)	37.8	2.21	0.31	0.09	45.7	2.43	0.17	0.09	46.6	1.79	0.12	0.58 <sup>*</sup>
13. Number you smoke influenced by other things (reversed)	64.6	1.41	-0.72	0.42	71.1	2.04	-0.67	0.15	72.1	1.52	-0.87	0.28
14. Number you smoke not affected much by other things	63.4	1.27	-0.71 <sup>a</sup>	0.33	72.7	1.12	-1.03	0.48	73.5	0.84	-1.40 <sup>d</sup>	0.71
15. Amount you smoke has increased	61.1	1.83 <sup>a</sup>	-0.48	0.13	66.8	3.04 <sup>a</sup>	-0.40	0.06	60.2	2.37	-0.32	0.17
16. Need to smoke a lot more to be satisfied.	50.7				60.8				56.5			
17. Can smoke much more now before you feel anything	47.6	2.02 <sup>a</sup>	-0.11	0.08	53.9	2.28 <sup>d</sup>	-0.29	0.09	49.0	1.87	-0.22	0.36 <sup>*</sup>

Note: a = Discrimination; b = Symptom Severity;

\* =  $p < .05$ ;

<sup>a</sup> = Values with the same superscript are significantly different across assessment waves.

**Table 3**  
IRT score to summed score conversion table for the DSM-IV, mFTQ, and NDSS.

	DSM-IV			mFTQ			NDSS		
	Summed Score	IRT Score	SD	Summed Score	IRT Score	SD	Summed Score	IRT Score	SD
0	-0.95	0.75	0.74	0	-1.62	0.74	0	-1.90	0.58
1	-0.39	0.68	0.66	1	-0.88	0.66	1	-1.53	0.51
2	0.12	0.62	0.65	2	-0.06	0.65	2	-1.20	0.45
3	0.58	0.58	0.63	3	0.49	0.63	3	-0.92	0.40
4	1.01	0.55	0.65	4	0.99	0.65	4	-0.68	0.37
5	1.43	0.54	0.66	5	1.46	0.66	5	-0.47	0.34
6	1.87	0.54	0.68	6	1.91	0.68	6	-0.27	0.33
7	2.39	0.56	0.67	7	2.22	0.67	7	-0.08	0.32
							8	0.12	0.32
							9	0.32	0.33
							10	0.53	0.35
							11	0.77	0.38
							12	1.06	0.43
							13	1.41	0.49
							14	1.84	0.57

DSM-IV: Diagnostic and Statistical Manual Fourth Edition criteria for Nicotine Dependence (7 symptoms); mFTQ: Modified Fagerstrom Tolerance Questionnaire (7 symptoms); NDSS: Nicotine Dependence Syndrome Scale 14 symptoms. All scale scores reflect a sum of dichotomized responses.