

An integrated data analysis approach to investigating measurement equivalence of DSM nicotine dependence symptoms

Jennifer S. Rose^{a,*}, Lisa C. Dierker^a, Donald Hedeker^b, Robin Mermelstein^b

^a Psychology Department, Wesleyan University, Middletown, CT 06459, United States

^b Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL 60608, United States

ARTICLE INFO

Article history:

Received 30 April 2012

Received in revised form 3 September 2012

Accepted 5 September 2012

Available online 27 September 2012

Keywords:

Nicotine dependence symptoms
Measurement equivalence
Integrative data analysis
Moderated nonlinear factor analysis
Adolescent

ABSTRACT

Introduction: Research identifying nicotine dependence (ND) symptoms most appropriate for measurement of adolescent ND and invariant across the range of smoking exposure is hampered by limited sample size and variability of smoking behavior within independent studies. Integrative data analysis, the process of pooling and analyzing data from multiple studies, produces larger and more heterogeneous samples with which to evaluate measurement equivalence across the full continuum of smoking quantity and frequency.

Methods: Data from two studies were pooled to obtain a large sample of adolescent and young adult smokers with considerable variability in smoking. We used moderated nonlinear factor analysis, which produces study equivalent ND scores, to simultaneously evaluate whether 14 DSM ND symptoms had equivalent psychometric properties (1) at different levels of smoking frequency and (2) across a continuous range of smoking quantity, after accounting for study differences.

Results: Nine of 14 symptoms were equivalent across levels of smoking frequency and quantity in probability of endorsement at different levels of ND and in ability to discriminate between levels of ND severity. A more precise ND factor score accounted for study and smoking related differences in symptom psychometric properties.

Conclusions: DSM-IV symptoms may be used to reliably assess ND in young populations across a wide range of smoking quantity and frequency and within both nationally representative and geographically restricted samples with different study designs. Symptoms shared across studies produced an equivalently scaled ND factor score, demonstrating that integrating data for the purpose of studying ND in young smokers is viable.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The most important factor believed to contribute to smoking persistence and failed cessation is nicotine dependence (ND). At the symptom level, research shows that some adolescents report features of ND (i.e., loss of control, craving, and tolerance) soon after smoking onset and at low levels of smoking (DiFranza et al., 2007a,b; Gervais et al., 2006; Rose et al., 2010). Thus, research aimed at elucidating the etiology of the addiction process must rely on measures that include symptoms that capture the full range of ND severity and discriminate among different levels of dependence.

Presently, the measurement of ND among adolescents is complicated by evidence of differential item functioning (DIF) showing systematic differences in symptom endorsement related to

adolescents' smoking frequency, independent of their actual level of ND severity (Rose and Dierker, 2010b). Ideally, individuals with the same level of ND should have the same likelihood of endorsing a symptom. That is, symptom psychometric properties should remain stable and independent of individual differences in smoking exposure, so that only true individual differences affect the probability of symptom endorsement. Otherwise, true individual differences in ND symptom endorsement are confounded with unwanted differences in the psychometric properties of the symptoms. This measurement bias can either inflate or weaken the association of ND with other important factors. For example, symptoms such as irritability and craving revealed DIF between nondaily and daily smokers such that daily smokers had a greater probability of endorsing these symptoms than nondaily smokers, despite having the same underlying level of ND severity as measured by the observed dependence symptoms. In other words, the probability of endorsement, which should only be influenced by individual differences in smoking behavior, was also being influenced by the symptoms' psychometric properties.

* Corresponding author at: Department of Psychology, Wesleyan University, 237 High Street, Middletown, CT 06459, United States. Tel.: +1 860 685 2406.

E-mail address: jrose01@wesleyan.edu (J.S. Rose).

There has been considerable psychometric research on DSM ND symptoms. This research indicates that symptoms are represented by a single underlying continuum of ND severity (Saha et al., 2010; Strong et al., 2009, 2012). These studies also suggest that DSM symptoms are best at measuring higher levels of ND (Rose and Dierker, 2010a; Strong et al., 2003, 2007), and that the psychometric properties of many of these symptoms appear equivalent across demographic characteristics such as age, gender and ethnicity (Saha et al., 2010; Strong et al., 2012), and for daily and nondaily adolescent smokers (Rose and Dierker, 2010a).

However, current research aimed at identifying symptoms most appropriate for measuring ND in adolescents that are invariant to the broad range of smoking exposure has been hampered by the fact that heavier smokers often are not adequately represented in individual studies of young smokers. This has made it difficult to examine measurement equivalence at more refined levels of smoking quantity and frequency. Instead, young smokers must be grouped more generally (e.g., nondaily and daily smokers), which does not solve the problem of underrepresentation of heavier smokers and ignores the broad range of lighter smoking levels typically found among adolescent smokers. These sample size issues also preclude an evaluation of measurement equivalence for multiple dimensions of smoking exposure (e.g., smoking frequency and quantity) simultaneously, forcing us to draw overly simplified conclusions about ND symptom equivalence across the full continuum of smoking exposure.

One potential solution would be to pool data from existing studies assessing adolescent and young adult smoking behavior and ND. Doing so would provide the sample sizes and variability in smoking behavior necessary to reliably assess ND symptom equivalence. Unfortunately, differences between studies in terms of target populations and measurement of ND, both of which serve as potential sources of measurement nonequivalence across studies, have prohibited us from pooling data from multiple studies. Recent innovations, however, in moderated nonlinear factor analysis (MNLFA) now make it possible to create measures that are equivalent across studies, allowing us to analyze integrated data from multiple studies despite differences in study design (Bauer and Hussong, 2009). Creating ND measures that share a common metric across studies would allow researchers to use the larger and more heterogeneous integrated data sets to address research questions that might not be adequately powered in separate independent studies.

Although creating equivalent measures across samples is well-established in the area of educational testing (c.f., Hambleton et al., 1995; Lord, 1980), equating ND measures from unrelated studies has not yet been undertaken. In addition, prior research on the psychometric properties of ND symptoms has used item response theory (IRT) modeling to evaluate symptom equivalence. One major advantage of MNLFA over IRT is that MNLFA permits an

examination of ND symptom equivalence across multiple categorical and continuous variables.

The purpose of this study was to extend existing research on the measurement of DSM ND symptoms to a larger population of adolescent and young adult smokers (ages 16–25) exhibiting a more heterogeneous range of smoking behavior than can be found in any one study, using a unique application of integrated data analysis (IDA) and MNLFA methods. Specifically, we sought to evaluate whether DSM ND symptoms were equivalent (1) across studies, (2) at different levels of smoking frequency, measured categorically, and (3) across a continuous range of smoking quantity.

2. Methods

2.1. Participants

Current smokers were drawn from two different studies (Table 1). First, the National Epidemiologic Study of Alcohol and Related Conditions (NESARC) used a three-stage sampling design to select a representative sample of US civilian, noninstitutionalized adults aged 18 and older. The NESARC oversampled Blacks, Hispanics, and individuals age 18–24. The present study included 1510 participants (age 18–25) from the 2001–2002 Wave 1 survey who reported smoking in the past 30 days. This sample was 48.6% female and 62% White, with a mean age of 21.6 years ($sd = 2.2$, range = 18–25). Eighty-one percent smoked daily (mean 10.9 cigarettes/day).

The Social and Emotional Contexts of Adolescent Smoking Patterns Study (SECASP) included adolescents at high risk for smoking. All ninth and tenth grade students at 16 Chicago-area high schools completed a brief screener survey of smoking behavior ($N = 12,970$). Students reporting smoking in the past 90 days and lifetime exposure of <100 cigarettes were invited to participate, as were those who smoked in the past 30 days and smoked >100 cigarettes in their lifetime. Additionally, random samples of youth who never smoked or smoked <100 lifetime, but not in the past 90 days, were also invited to participate. DSM ND symptoms were measured for the first time among past month smokers at the 24 month follow-up in 2007–2008 ($N = 488$). This 24 month sample was 54.5% female with a mean age of 17.7 years (range = 16–19). The majority 66% were White. Only 24% smoked daily (mean 3.9 cigarettes/day).

2.1.1. Pooled sample. The final data set was created by pooling the two samples described above ($N = 1998$ participants). The combined sample exhibited considerably more heterogeneity in age and smoking frequency and quantity compared to either the NESARC or SECASP studies alone (Table 1).

2.2. Measures

2.2.1. Smoking. In the pooled sample, current smoking frequency was categorized into three levels representing less than weekly, at least once a week, and daily smoking. A continuous measure of current smoking quantity was created for the pooled sample. In the NESARC, participants reported the actual number of cigarettes they currently smoked per day. The SECASP had response options representing ranges of daily smoking quantity. Therefore, for the SECASP study, we imputed the midpoint of the response options that included a range (e.g., 8 for response option 6–10 cigarettes/day) and recoded >20 cigarettes/day to 25 in both studies.

2.2.2. Nicotine dependence. Items representing DSM ND symptoms are shown in Table 2. Differences across studies included response scales, time frame for response, symptom types, phrasing of questions, and number of items used to measure some symptoms. The NESARC used the Alcohol Use Disorder and Associated Disabilities

Table 1
NESARC, SECASP and pooled sample characteristics.

Characteristic	NESARC ($N = 1510$)	SECASP ($N = 488$)	Pooled sample ($N = 1998$)
Female*	48.6%	54.5%	50.1%
Mean age (range)*	21.6 (18–25)	17.7 (16–19)	20.6 (16–25)
Ethnicity			
White	62.2%	65.9%	63.1%
Black	13.1%	10.6%	12.5%
Hispanic	18.5%	19.1%	18.6%
Other	6.2%	4.3%	5.8%
Smoking frequency*			
Less than weekly	5.0%	40.4%	13.7%
Weekly or more	13.9%	34.8%	19.0%
Daily	81.1%	24.8%	67.3%
Mean smoking quantity* (sd)	10.9 (7.2)	3.9 (4.2)	9.2 (7.2)
Mean age of smoking onset* (sd)	14.9 (2.9)	12.8 (2.2)	14.4 (2.9)

* Significantly different between studies ($p < .05$).

Table 2
Nicotine dependence symptoms from the NESARC and SECASP studies.

Symptom	Assessed in NESARC	Assessed in 24 month SECASP
Withdrawal	Yes	Yes
Withdrawal symptoms uncomfortable/upsetting/interfere w/work, social activities	Yes	Yes ^a
Smoke to keep from feeling withdrawal symptoms	Yes	Yes
Smoke despite promising yourself not to	No	Yes
Use more frequently than intended	Yes	Yes
Tried to cut down/quit but could not	Yes	Yes
Chain smoking	Yes	Yes
Give up activities to use	Yes ^a	Yes
Health problem	Yes	Yes
Emotional problem	Yes	Yes
Strong desire to smoke	No	Yes
Need to smoke more to feel satisfied	No	Yes
Need to smoke more to feel effects	Yes	Yes
Increased use by 50% or more	Yes	No

^a Two separate questions were combined to create a single question to be consistent with a single question in the other study.

Interview Schedule-IV (AUDADIS-IV; Grant et al., 2001) tobacco module to assess whether participants had experienced 18 ND symptoms in the past 12 months using a binary yes/no response scale. Of the 18 NESARC symptoms, 8 represented withdrawal (e.g., “irritability”, “difficulty concentrating”, “depressed mood” after not smoking for a while). These symptoms were combined into a single withdrawal symptom using the DSM definition of 4 or more endorsed symptoms to indicate positive endorsement of withdrawal.

The 24-month follow-up of the SECASP study assessed 22 DSM symptoms using the TTURC ND inventory (Dierker et al., 2006; Sledjeski et al., 2007). Participants were asked to indicate whether they experienced symptoms in the past 30 days on a 4-point response scale ranging from 1 = not at all to 4 = quite a bit. Because MNLFA requires that symptoms that are common to both studies have the same response scale and because of low endorsement rates for the “somewhat” and “quite a bit” categories, we dichotomized the SECASP responses into two categories reflecting no symptom endorsement (0 = not at all) vs. any endorsement (1 = any of the other three responses). Ten of the SECASP symptoms represented withdrawal, and were combined into a single withdrawal symptom using the DSM definition described above. Of the 14 pooled symptoms, 1 was unique to the NESARC and 3 were unique to the SECASP.

2.2.3. Study membership. Study membership was a dichotomous variable coded 0 for participants in the NESARC sample and 1 for participants in the SECASP study.

2.3. Analysis

We used MNLFA to evaluate the extent to which severity (probability of endorsement at different levels of the underlying ND construct) and discrimination (ability of each symptom to discriminate between different levels of the underlying ND construct) differed for each ND symptom across studies and levels of smoking frequency and quantity in the pooled data set. MNLFA does not require that all symptoms are the same across studies, resulting in a larger pool of symptoms than typically available in independent data sets. Although study DIF cannot be tested for symptoms that are not shared across studies, smoking frequency and quantity DIF can be tested. Preliminary exploratory factor analyses to determine the number of factors to test in the MNLFA models supported a single underlying ND construct in both samples, which is consistent with past research (Courvoisier and Etter, 2008; Etter et al., 1999; Rose and Dierker, 2010a; Strong et al., 2009).

The MNLFA analyses proceeded in several steps outlined by Bauer and Hussong (2009). First, we tested a nonlinear 1-factor model consisting of the 14 binary symptoms without moderators, thus assuming no symptom DIF. The mean and variance of the latent factor were fixed to 0 and 1, respectively, to identify the model and obtain parameter estimates for all symptom intercepts and slopes. This model served as the baseline model to which subsequent models with moderators were compared via the likelihood ratio (LRT) chi-square test to determine whether allowing DIF significantly improved the fit of the models. Second, we tested a model that allowed the factor mean and variance to vary as a function of study membership, smoking frequency and smoking quantity because differences in the factor mean and variance related to the moderators can appear as DIF in the individual symptoms. Third, we tested models allowing symptom parameters (intercept and slope) to vary as a function of smoking quantity, frequency and study membership, one symptom at a time. LRT chi-square tests comparing these models to the baseline model indicated whether there was significant DIF. For symptoms showing DIF, we examined significance tests for intercept and slope parameters to determine whether there was significant DIF for the intercept, slope, or both. Finally, we tested a model that allowed factor moderation and DIF for all symptom parameters showing significant DIF in the previous step. All MNLFA models were analyzed using SAS PROC NLMIXED (SAS Institute, 2008), using the Benjamini–Hochberg method to correct for Type I error inflation (Benjamini and Hochberg, 1995). Type I error adjustments were made

first for the LRT chi-square tests comparing each symptom DIF model to the baseline model and then again for significance tests of the final model DIF parameter estimates.

The final model was used to generate item characteristic curves (ICCs) to show estimated probabilities of endorsement for each symptom at levels of the ND construct and to calculate a modal a posteriori (MAP) ND severity factor score. The MAP score represents the mode of the joint likelihood of the product of the population distribution and the MNLFA parameter estimates. Unlike traditional ND scoring methods, which assume equal weighting of each symptom on the underlying ND construct, the MAP score takes into account each individual’s response pattern, individual symptom severity and discrimination estimates, and DIF as a function of study and smoking frequency and quantity differences, providing a standard normal ND score (mean = 0, sd = 1) for each participant in the pooled sample. Theoretically, to create an ND score equivalent across studies, at least one symptom must show no study DIF across studies, but ND score reliability and confidence in the common metric increases as the number of equivalent symptoms increases. Researchers in the area of educational test equating recommend that at least 20% of the items be equivalent (Kolen and Brennan, 2004). If all symptoms show study DIF, then IDA for the purpose of studying ND cannot proceed (Bauer and Hussong, 2009). The distribution of this factor score was compared to the distribution of a score representing the proportion of endorsed symptoms.

3. Results

3.1. Symptom prevalence rates

Table 3 shows symptom endorsement rates for each sample as well as intercept (severity) and factor loading (discrimination) parameter estimates for the final MNLFA model. Symptoms most prevalent in the NESARC study included “health problem” related to smoking, “tried to cut down/quit but could not”, and “chain smoking”. Symptoms most prevalent in the SECASP study included “needing to smoke more to feel effects”, “smoking despite promising yourself not to”, and “needing to smoke more to feel satisfied”.

3.2. Moderation of symptom severity (intercepts) and discrimination (slopes)

Of the 14 symptoms, 9 showed no significant DIF related to study membership, smoking quantity or smoking frequency despite considerable differences in samples and ND measurement across studies and levels of smoking. Symptom ICCs were generally steep indicating good discrimination between levels of ND (Fig. 1), with symptoms related to withdrawal discriminating best. Discrimination was lowest for “needing to smoke more to feel satisfied”, “giving up activities to use”, and “increased use by 50% or more”. The curves were also shifted to the right of the mean ND factor score, indicating that symptoms were most likely to be endorsed at higher levels of ND. The symptoms “giving up activities to use” and “increased use by 50%” had the highest severity. Symptoms “need to smoke more to feel satisfied” and “use more frequently

Table 3
MNLFA Estimated symptom psychometric properties.^a

Symptom	Observed probability of endorsement		Intercept	Loading	Estimated probability of endorsement for ND severity score = 1
	NESARC	SECASP			
Withdrawal ^b	.18	.49	−10.78	5.52	.01
Withdrawal symptoms uncomfortable/upsetting/interfere w/work, social activities ^b	.11	.47	−8.80	4.42	.01
Smoke to keep from feeling withdrawal symptoms ^b	.18	.51	−8.25	4.19	.02
Smoke despite promising yourself not to	NA	.60	−2.61 ^{LF} −2.33 ^{WF} −3.88 ^{DF}	1.78	.30 ^{LF} .37 ^{WF} .11 ^{DF}
Use more frequently than intended ^b	.31	.58	−3.02	1.92	.25
Tried to cut down/quit but could not	.44	.47	−3.48 ^N −5.30 ^S	2.57	.29 ^N .06 ^S
Chain smoking ^b	.32	.37	−4.15 ^{LQ} −3.37 ^{MQ} −2.58 ^{HQ}	1.73	.08 ^{LQ} .16 ^{MQ} .30 ^{HQ}
Give up activities to use ^b	.07	.17	−5.03	1.59	.04
Health problem	.45	.44	−1.88 ^{N,LQ} −1.2 ^{N,MQ} −0.5 ^{N,HQ} −3.00 ^{S,LQ} −2.33 ^{S,MQ} −1.65 ^{S,HQ}	0.87 ^{LF} 1.48 ^{WF} 1.84 ^{DF}	.27 ^{N,LQ,LF} .41 ^{N,MQ,LF} .58 ^{N,HQ,LF} .49 ^{N,LQ,DF} .65 ^{N,MQ,DF} .79 ^{N,HQ,DF} .11 ^{S,LQ,LF} .19 ^{S,MQ,LF} .31 ^{S,HQ,LF} .24 ^{S,LQ,DF} .38 ^{S,MQ,DF} .55 ^{S,HQ,DF}
Emotional problem ^b	.14	.35	−5.52	2.39	.04
Strong desire to smoke	NA	.41	−6.97	3.21	.02
Need to smoke more to feel satisfied ^b	.14	.59	−1.29	1.37	.52
Need to smoke more to feel effects	NA	.72	−3.48 ^{LF} −7.00 ^{WF} −6.05 ^{DF}	2.04	.19 ^{LF} .02 ^{WF} .01 ^{DF}
Increased use by 50% or more	.15	NA	−4.25	1.69	.07

Note. In the mixed model parameterization, higher (more positive) intercepts reflect lower severity, and higher loadings reflect greater discrimination. NA = not measured. Superscripts indicate parameter estimates for study (N = NESARC; S = SECASP), smoking quantity (LQ = low quantity (−1 sd), MQ = mean quantity, HQ = high quantity (+1 sd)), and smoking frequency (LF = less than weekly smoking, WF = weekly, DF = daily smoking).

^a Using the Benjamini–Hochberg correction for multiple significance testing.

^b Significantly different ($p < .05$) symptom prevalence rates between studies based on bivariate chi-square tests.

than intended” had the lowest severity, indicating that they were more likely to be endorsed at lower levels of ND severity.

The remaining five symptoms showed DIF on at least one of the moderators. Less than weekly smokers were significantly more likely to endorse “smoke despite promising not to” at lower levels of ND severity compared to daily smokers (DIF estimate = −1.26, $z = -3.13$, $p = .0018$). Less than weekly smokers were also more

likely to endorse “needing to smoke more to feel effects” compared to both weekly (DIF estimate = −3.52, $z = -3.13$, $p = .0018$) and daily smokers (DIF estimate = −2.57, $z = -3.49$, $p = .0005$) (Fig. 2). The symptom “tried to cut down/quit” (DIF estimate = −1.82, $z = -8.22$, $p < .0001$) was more likely to be endorsed at lower levels of severity for NESARC compared to SECASP participants (Fig. 3). The symptom “chain smoking” (DIF estimate = .085, $z = 9.88$, $p < .0001$) was

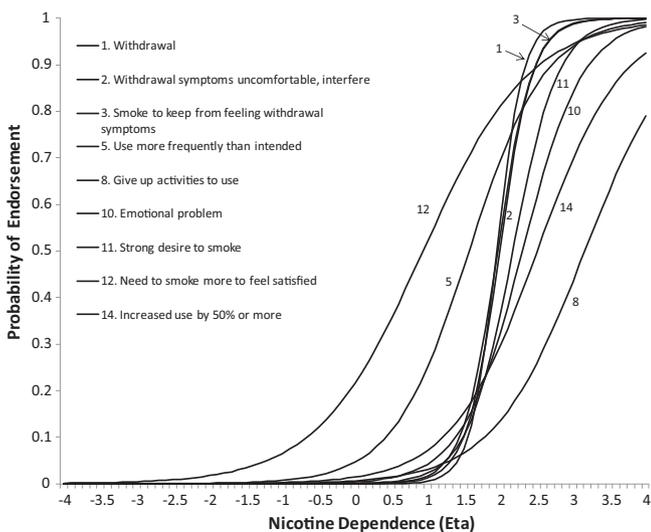


Fig. 1. ICCs for symptoms without DIF.

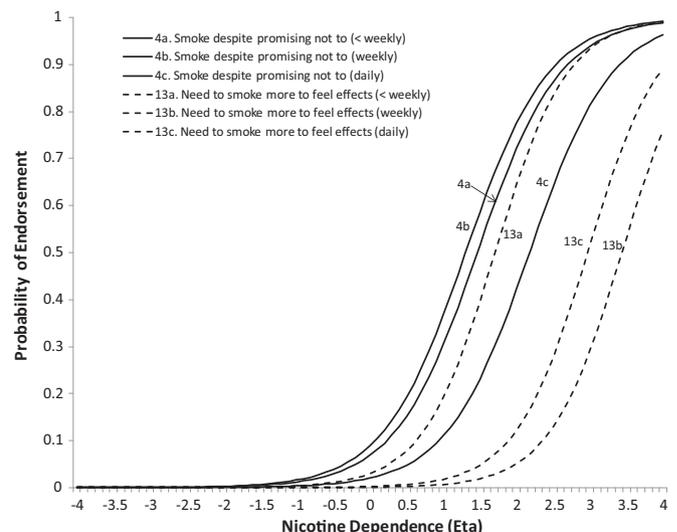


Fig. 2. ICCs showing smoking frequency DIF for the smoke despite promising not to and need to smoke more to feel effects symptoms.

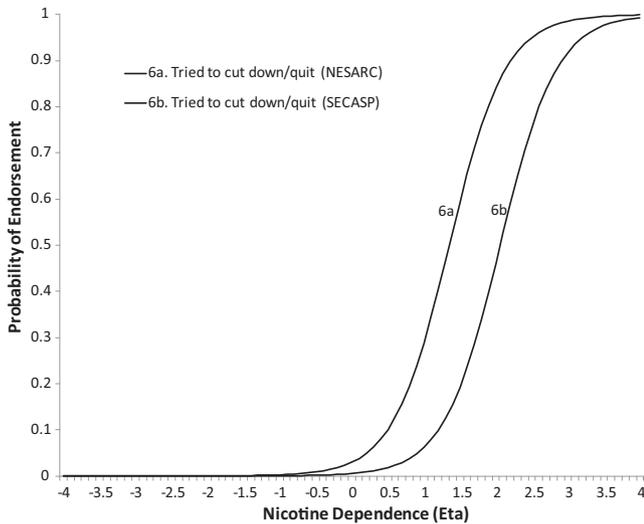


Fig. 3. ICC showing study related DIF for the tried to quit/cut down but could not symptom.

more likely to be endorsed at lower levels of ND severity for those with greater smoking quantity (Fig. 4). Finally, the symptom “health problem” was more likely to be endorsed at lower levels of ND for NESARC participants (DIF estimate = -1.11, $z = -5.08$, $p < .0001$) and for those with greater smoking quantity (DIF estimate = -.073, $z = 2.09$, $p = .0038$), and showed lower discrimination for those smoking less than weekly compared to daily smokers. Figs. 5 and 6 (broken down by smoking frequency for ease of presentation) show the ICCs for DIF on all three moderators.

The magnitude of DIF was assessed by examining difference between groups in the probability of symptom endorsement at a given level of ND severity. For example, the DIF estimate for “smoke despite promising not to” suggested that, for a standardized ND severity score of 1, the probability of endorsement was .30 for less than weekly smokers compared to only .11 for daily smokers. This is a substantial difference given that, if the underlying ND factor score is the same, the probability of symptom endorsement should be the same regardless of how often individuals smoke. Differences of this magnitude were observed for the other DIF parameters as well (Table 3).

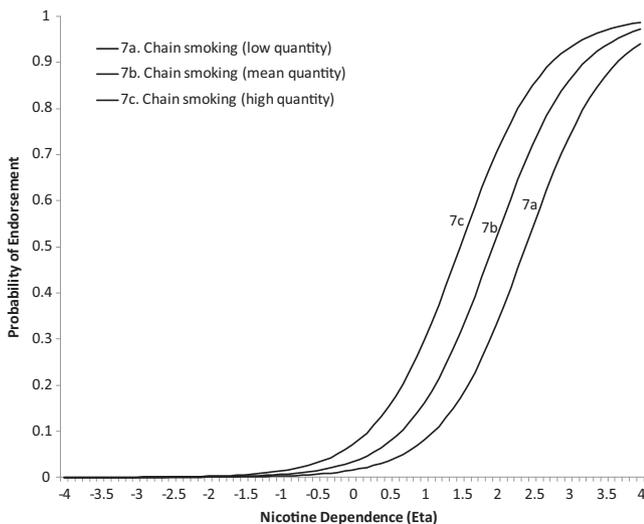


Fig. 4. ICC showing smoking quantity DIF for chain smoking.

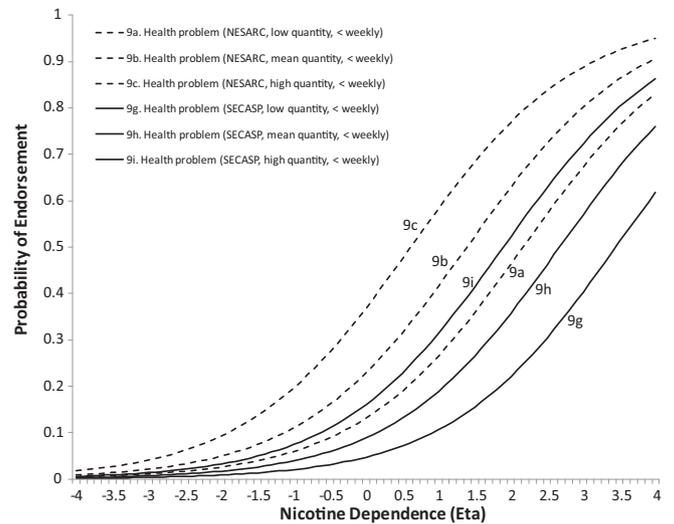


Fig. 5. ICCs showing study and smoking quantity DIF for the health problem symptom for less than weekly smokers.

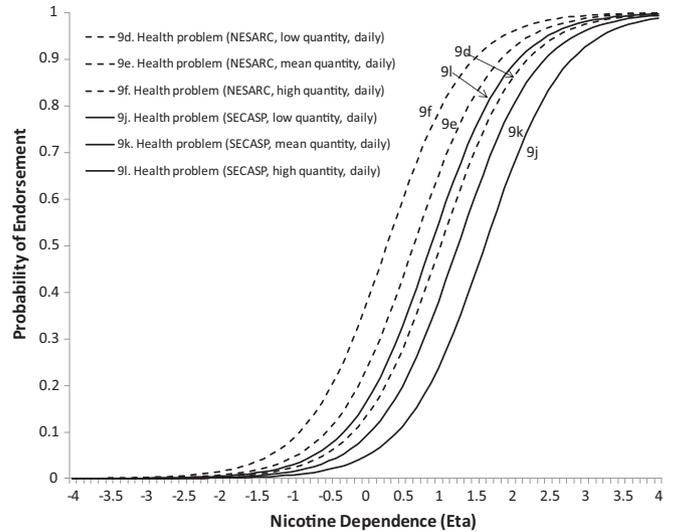


Fig. 6. ICCs showing study and smoking quantity DIF for the health problem symptom for daily smokers.

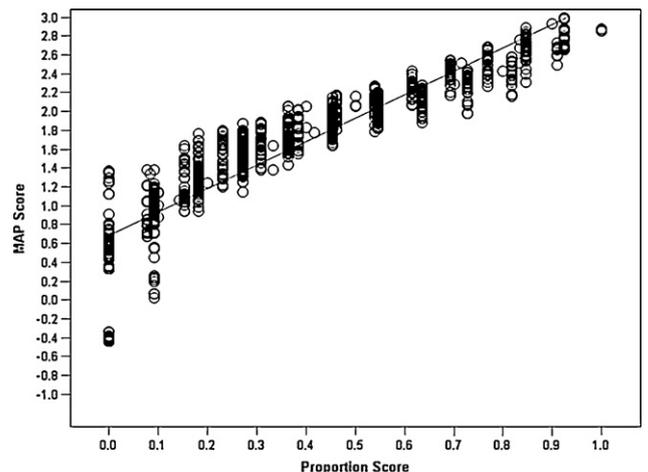


Fig. 7. Scatterplot of the ND symptom proportion scores and MAP scores.

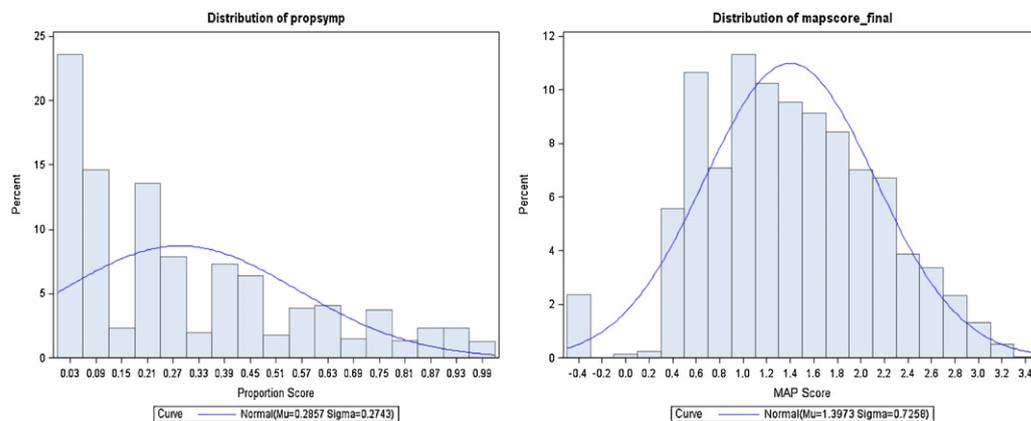


Fig. 8. Distribution of ND symptom proportion score and MAP score.

3.3. Comparison of MNLFA MAP factor score and proportion of endorsed symptoms score

Fig. 7 shows a scatterplot of the ND MAP score and the proportion score. Although the two scores were highly correlated ($r = .94$), which suggests a high level of agreement in terms of relative rank (i.e., increases in proportion scores were closely related to increases in the MAP score), there was considerably more variability in MAP factor scores at each level of the proportion score. For example, given a single proportion score of .50, there were 28 different MAP scores, suggesting greater precision in scoring ND for the MAP score (Fig. 8).

4. Discussion

This study is the first to investigate ND symptom measurement equivalence using data pooled from two *unrelated* studies of young smokers with a wide range of smoking behavior. Advantages of integrative data analysis (IDA) include replication of symptom psychometric properties across multiple studies using a larger pool of symptoms, increased sample size and heterogeneity, all permitting examination of symptom properties across studies and a full range of smoking exposure. Thus, we were able to conduct a broader, more powerful and generalizable, psychometric assessment of ND symptom properties, and could evaluate these properties over and above differences in how they were measured.

In general, the DSM symptoms assessed in this study showed remarkable stability across studies and levels of smoking frequency and quantity. A total of 9 of the 14 symptoms were found to be equivalent, including the 3 symptoms reflecting withdrawal, 2 of 3 symptoms reflecting tolerance (“need to smoke more to feel satisfied” and “increased use”), craving (“strong desire to smoke”), “giving up activities to smoke”, “using more often than intended”, and “continuing to smoke despite emotional problems”. These findings are rather remarkable because they indicate that symptom psychometric properties were not influenced by extensive differences in study design, or by level of smoking exposure. Researchers and clinicians wanting robust symptoms to measure ND without calculating factor scores might consider including these symptoms, either individually or collapsed into DSM criteria.

The majority of symptoms also discriminated well between levels of ND. Equivalent symptoms of “giving up activities to smoke”, “increased use by 50% or more”, and “needing to smoke more to feel satisfied” had the lowest discrimination relative to the other symptoms, yet remained reasonable. Only one symptom, “health problem”, had a nonequivalent discrimination parameter, indicating that this symptom was more sensitive to changes in ND severity for more frequent smokers. These results replicated past research

showing that DSM criteria measured the ND construct well and were equivalent across levels of smoking frequency in a nationally representative sample (Rose and Dierker, 2010a), and extended this research to show that equivalence is also present in a more geographically restricted, nonrandom sample that used a different sampling strategy.

All five nonequivalent symptoms showed symptom severity DIF. Two showed study DIF, two showed smoking quantity DIF, and two showed smoking frequency DIF. Less frequent smokers were more likely to endorse “smoke despite promising not to” and “need to smoke more to feel effects” at lower levels of severity, whereas NESARC participants and more frequent, heavier smokers were more likely to endorse “tried to quit/cut down”, “chain smoking”, and “health problem” at lower levels of severity. This might be due to differences in the salience of symptoms at different stages of smoking. For less frequent smokers who may just be beginning to show symptoms, “smoke despite promising not to” and “need to smoke more to feel effects” may be more salient, whereas symptoms like “chain smoking”, “tried to quit/cut down”, and “health problem” are likely to be much more salient to heavier smokers. It is not particularly difficult to use MNLFA to obtain factor scores using SAS, a commercially available software package. Researchers might consider using MNLFA to identify DIF and create ND factor scores, or alternatively, using a subset of equivalent ND symptoms to measure ND severity. The “health problem” symptom was particularly unstable, indicating that it may not be ideal for use with young smokers with lower levels of smoking exposure.

The advantages of using integrated data are especially clear for investigating smoking related DIF. Investigating smoking frequency and quantity DIF in the SECASP alone would not have captured the full continuum of smoking behavior, as these participants were largely infrequent, low quantity smokers. Conversely, examining smoking frequency and quantity DIF in the NESARC alone would not have allowed us to determine whether DIF was evident at lower levels of smoking frequency and quantity. Pooling the samples, however, increased the sample size and expanded the range of smoking behavior for which we could assess ND symptom DIF, allowing us to generalize across a much broader range of smoking.

The presence of a large number of symptoms whose psychometric properties did not vary across studies allowed ND scores to be scaled equivalently across studies, thus breaking down a major barrier to IDA. Taking advantage of information provided by the unique characteristics of both studies also served to increase ND factor score reliability. Finally, compared to a simpler ND proportion score, the interval-scaled and normally distributed ND factor score showed more precision by capturing greater variability in ND.

Major strengths of this study include further evidence supporting the strong psychometric properties of DSM symptoms for use

with young smokers with a wide range of smoking frequency and quantity and illustration of the viability of IDA in adolescent ND research. Although we could identify study related DIF, we could not differentiate which aspects of the studies were responsible for the lack of equivalence. Ideally, a large number of studies could be pooled, and a multilevel analytic approach could be used to directly model the influence of specific study characteristics on symptom equivalence (Curran and Hussong, 2009). In addition, it is not known whether the strengths of an ND factor score translate into significant gains in ability to predict future smoking behavior. It is easier to create a score that is simply a function of the number of symptoms endorsed, so the cost/benefit ratio of generating ND factor scores needs to be evaluated. However, a simpler ND score does not guarantee measurement equivalence required for IDA, nor does it account for smoking related DIF.

In sum, this study used a novel application of IDA and MNLFA to evaluate DSM ND symptom measurement equivalence in a pooled sample of young smokers drawn from two studies with substantially different populations, sampling strategies, and ways of measuring the symptoms. Results suggest that DSM ND symptoms may be used to reliably assess ND in populations of young smokers with a wide range of smoking frequency and quantity. More importantly, this study showed that IDA to assess ND in young smokers is viable, making it possible to test hypotheses that previously could not be tested due to limitations inherent to independent studies. For example, pooling observations from multiple studies could lead to sample sizes large enough to investigate ND in typically under-represented ethnic groups. Future studies also may integrate data sets that use entirely different measures tapping similar dimensions of ND, such as the NDSS (Shiffman et al., 2004), to identify symptoms that are equivalent across measures.

Some evaluations of cross-study equivalence, particularly across different measures of ND, may fail to identify any study equivalent items. A collaborative approach to the design of future studies of smoking and ND in adolescent and young adults may allow us to take better advantage of the power of IDA across a larger number of independent studies and measures (Hofer and Piccinin, 2009). Regardless, IDA remains a promising methodology to address the extent to which different measures of ND assess the same construct; to continue to develop a pool of symptoms that reliably assess ND for young smokers with varying smoking exposure across a wide range of studies; and to identify and account for sources of nonequivalence.

Role of funding source

This research was supported by Project Grant P01CA098262 (Mermelstein) from the National Cancer Institute, R01 DA022313 A2, R01 DA022313 S1 (Dierker), and R21 DA029834-01 (Rose) from the National Institute on Drug Abuse, and Center Grant P50 DA010075 awarded to Penn State University. The content of this project is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. These agencies had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Contributors

All authors participated in the preparation of this manuscript. Authors Jennifer Rose and Lisa Dierker designed the study and wrote the first draft of the manuscript. Author Jennifer Rose undertook the statistical analysis. Dr. Hedeker assisted in the development of the data analytic plan, provided consultation on the

analyses, and reviewed manuscript drafts. Dr. Mermelstein provided consultation on the design of the study, provided feedback on statistical results, and reviewed drafts of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interest

All other authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by Project Grant P01CA098262 (Mermelstein) from the National Cancer Institute, R01 DA022313 A2, R01 DA022313 S1 (Dierker), and R21 DA029834-01 (Rose) from the National Institute on Drug Abuse, and Center Grant P50 DA010075 awarded to Penn State University. The content of this project is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

References

- Bauer, D.J., Hussong, A.M., 2009. Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychol. Methods* 14, 101–125.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple significance testing. *J. Roy. Stat. Soc. B* 57, 289–300.
- Courvoisier, D., Etter, J.-F., 2008. Using item response theory to study the convergent and discriminant validity of three questionnaires measuring cigarette dependence. *Psychol. Addict. Behav.* 22, 391–401.
- Curran, P.J., Hussong, A.M., 2009. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychol. Methods* 14, 81–100.
- Dierker, L., Richardson, E., Stolar, M., Flay, B., Tiffany, S., Collins, L., Bailey, S., Nichter, M., Nichter, M., Clayton, R., Tobacco Etiology Research Network (TERN), 2006. The proximal association between smoking and drinking among first year college students. *Drug Alcohol Depend.* 81, 1–9.
- DiFranza, J.R., Savageau, J.A., Fletcher, K., O'Loughlin, J., Pbert, L., Ockene, J.K., McNeill, A.D., Hazelton, J., Friedman, K., Dussault, G., Wood, C., Wellman, R.J., 2007a. Symptoms of tobacco dependence after brief intermittent use: the development and assessment of nicotine dependence in youth-2 study. *Arch. Pediatr. Adolesc. Med.* 61, 704–710.
- DiFranza, J.R., Savageau, J.A., Fletcher, K., Pbert, L., O'Loughlin, J., McNeill, A.D., Ockene, J.K., Friedman, K., Hazelton, J., Wood, C., Dussault, G., Wellman, R.J., 2007b. Susceptibility to nicotine dependence: the development and assessment of nicotine dependence in youth 2 study. *Pediatrics* 120, e974–e983.
- Etter, J.-F., Vu Duc, T., Perneger, T.V., 1999. Validity of the Fagerstrom test for nicotine dependence and of the heaviness of smoking index among relatively light smokers. *Addiction* 94, 269–281.
- Gervais, A., O'Loughlin, J., Meshefedjian, G., 2006. Milestones in the natural course of onset of cigarette use among adolescents. *CMAJ* 175, 255–261.
- Grant, B., Dawson, D., Hasin, D., 2001. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version. NIH, Bethesda, MD.
- Hambleton, R.K., Swaminathan, R., Rogers, J.E., 1995. *Item Response Theory: Principles and Applications*. John Wiley & Sons, New York.
- Hofer, S.M., Piccinin, A.M., 2009. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychol. Methods* 14, 150–164.
- Kolen, M.J., Brennan, R.L., 2004. *Test Equating Methods and Practices*, second ed. Springer, New York.
- Lord, F.M., 1980. *Applications of Item Response Theory to Practical Testing Problems*. Erlbaum, New Jersey.
- Rose, J.S., Dierker, L.C., 2010a. DSM-IV nicotine dependence symptom characteristics for recent-onset smokers. *Nicotine Tob. Res.* 12, 278–286.
- Rose, J.S., Dierker, L.C., 2010b. An item response theory analysis of nicotine dependence symptoms in recent onset adolescent smokers. *Drug Alcohol Depend.* 110, 70–79.
- Rose, J., Dierker, L., Donny, E., 2010. Nicotine dependence symptoms among recent onset adolescent smokers. *Drug Alcohol Depend.* 106, 126–132.
- Saha, T.D., Compton, W.M., Pulay, A.J., Stinson, F.S., Ruan, W.J., Smith, S., Grant, M.B.F., 2010. Dimensionality of DSM-IV nicotine dependence in a national sample: an item response theory application. *Drug Alcohol Depend.* 108, 21–28.
- SAS Institute, 2008. *SAS/STAT 9.2 User's Guide*. SAS Institute, Inc., Cary, NC.
- Shiffman, S., Waters, A.J., Hickcox, M., 2004. The Nicotine Dependence Syndrome Scale: a multidimensional measure of nicotine dependence. *Nicotine Tob. Res.* 6, 327–348.
- Sledjeski, E.M., Dierker, L.C., Costello, D., Shiffman, S., Donny, E., Flay, B.R., 2007. Predictive validity of four nicotine dependence measures in a college sample. *Drug Alcohol Depend.* 87, 10–19.

- Strong, D.R., Kahler, C.W., Ramsey, S.E., Brown, R.A., 2003. Finding order in the DSM-IV nicotine dependence syndrome: a Rasch analysis. *Drug Alcohol Depend.* 72, 151–162.
- Strong, D.R., Kahler, C.W., Abrantes, A.M., MacPherson, L., Myers, M.G., Ramsey, S.E., Brown, R.A., 2007. Nicotine dependence symptoms among adolescents with psychiatric disorders: using a Rasch model to evaluate symptom expression across time. *Nicotine Tob. Res.* 9, 557–569.
- Strong, D.R., Kahler, C.W., Colby, S.M., Griesler, P.C., Kandel, D., 2009. Linking measures of adolescent nicotine dependence to a common latent continuum. *Drug Alcohol Depend.* 99, 296–308.
- Strong, D.R., Schonbrun, Y.C., Schaffran, C., Griesler, P.C., Kandel, D., 2012. Linking measures of adult nicotine dependence to a common latent continuum and a comparison with adolescent patterns. *Drug Alcohol Depend.* 120, 88–98.