Psychological Assessment
2009, Vol. 21, No. 4, 595–607

1040-3590/09/$12.00 DOI: 10.1037/a0017312

Do Smokers Know What We're Talking About? The Construct Validity of Nicotine Dependence Questionnaire Measures

Sandra J. Japuntich, Megan E. Piper, and Tanya R. Schlam
University of Wisconsin School of Medicine and Public Health

Daniel M. Bolt
University of Wisconsin—Madison

Timothy B. Baker
University of Wisconsin School of Medicine and Public Health

Few studies have examined whether nicotine dependence self-report questionnaires can predict specific behaviors and symptoms at specific points in time. The present study used data from a randomized clinical trial ($N = 608$; M. E. Piper et al., 2007) to assess the construct validity of scales and items from 3 nicotine dependence measures: the Fagerström Test for Nicotine Dependence (T. F. Heatherton, L. T. Kozlowski, R. C. Frecker & K.-O. Fagerström, 1991), the Nicotine Dependence Syndrome Scale (S. Shiffman, A. J. Waters, & M. Hickcox, 2004), and the Wisconsin Inventory of Smoking Dependence Motives (M. E. Piper et al., 2004). Scales from these measures were used to predict participants' reports on real-time measures of withdrawal symptoms and smoking behavior and retrospective self-report questionnaires to assess convergent and discriminative validity. The nicotine dependence measures' scales and items generally predicted the real-time measures of similar constructs, but the percent of variance accounted for was low. The nicotine dependence measures did, however, show evidence of discriminative validity. Thus, this study provides modest support for the construct validity of these nicotine dependence scales.

Keywords: nicotine dependence, construct validity, self-report, cigarette smoking, ecological momentary assessment

In the field of tobacco dependence, self-report questionnaires are used extensively in both clinical practice and research. Clinicians use self-report items to set dosing levels for nicotine replacement therapy (e.g., nicotine lozenge dose is based on the latency between waking and reported time of smoking the first cigarette of the day; Item 1 of the Fagerström Test for Nicotine Dependence; Heatherton, Kozlowski, Frecker, & Fagerström, 1991). Researchers administer self-report measures to assess the severity of nicotine dependence, and they use resulting estimates to test the heritability of nicotine dependence and to estimate the likelihood of cessation success (e.g., Baker et al., 2007, 2009; Beirut et al., 2007; Cannon et al., 2005; Etter, Le Houezec, & Perneger, 2003; Haberstick et al., 2007; Hyland et al., 2006; Kozlowski, Porter, Orleans, Pope, & Heatherton, 1994; Saccone et al., 2007). In addition, researchers use self-report measures to test assumptions about the structure of dependence such as whether dependence is multidimensional and whether there are qualitatively different subtypes of dependence (e.g., Goedeker & Tiffany, 2008; Hudmon, Gritz, Clayton, & Nisenbaum, 1999; Muthén & Asparouhov, 2006; Piper et al., 2008; Shiffman, Waters, & Hickcox, 2004). Researchers also use self-report measures to make inferences about the nature of smoking motives (e.g., Piper et al., 2004), and participants' responses are assumed to reflect their actual smoking behavior (i.e., responses are often interpreted as though they have face validity). However, the construct validity of such items is rarely comprehensively assessed.

The word *construct*, as in construct validity, can be defined as "some postulated attribute of people, assumed to be reflected in test performance" (Cronbach & Meehl, 1955, p. 283). It is not possible to measure a construct directly, and a construct cannot be represented fully by a single variable. Rather, a construct is defined by the interlocking system of laws that relates the construct to other constructs and to observable elements or manipulations of the environment (Wiggins, 1973). This interlocking system of laws—the nomological network (Cronbach & Meehl, 1955)—connects observable variables and constructs through theory-based lawful relations. Thus, in creating a measure of a construct, the investigator posits a pattern of associations that *should* exist if the measure is indeed assessing the targeted construct. *Construct validation*, then, refers to the process of examining whether the

Sandra J. Japuntich, Megan E. Piper, and Tanya R. Schlam, and Timothy B. Baker, Center for Tobacco Research and Intervention, University of Wisconsin School of Medicine and Public Health; Daniel M. Bolt, Department of Educational Psychology, University of Wisconsin—Madison.

Timothy B. Baker has served as an investigator on research projects sponsored by pharmaceutical companies including Pfizer, Glaxo Wellcome, Sanofi, and Nabi.

This research was supported by National Institutes of Health Grants CA84724-05, DA0197-06, 1K05CA139871, and 1KL2RR025012-01.

Correspondence concerning this article should be addressed to Sandra J. Japuntich, who is now at the Institute for Health Policy, 50 Staniford Street, 9th Floor, Boston, MA 02114. E-mail: sjapuntich@partners.org

measure of a construct actually has a network of associations that confers meaning. Construct validity is established if a measure correlates strongly with variables with which it is purported to be associated, and is less strongly related to other variables (Campbell & Fiske, 1959).[1] In the present study, we assessed the construct validity of both scales and individual items from three commonly used nicotine dependence self-report questionnaires: the Fagerström Test for Nicotine Dependence (FTND; Heatherton et al., 1991), the Nicotine Dependence Syndrome Scale (NDSS; Shiffman et al., 2004), and the Wisconsin Inventory of Smoking Dependence Motives (WISDM; Piper et al., 2004).

Construct validity has several components that represent diverse sorts of relations among variables. According to Campbell and Fiske (1959), the construct validation of a measure can be inferred on the basis of a series of comparisons of (a) measures of the same construct using different methods and (b) measures of a different construct using the same method. They label these comparisons multitrait–multimethod (MTMM) comparisons. These comparisons are described by Campbell and Fiske (1959) as falling into two major categories: convergent validity and discriminant validity.[2] Convergent validity is the level of agreement between a measure and other measures of the same construct. Strong evidence of convergent validity can be found when there is a high correlation between measures of the same construct, using different methods. For example, a new nicotine dependence measure might acquire convergent validation to the extent that it correlates positively with an existing dependence questionnaire or with a behavioral manifestation of dependence (e.g., cigarette consumption). *Discriminant validity* refers to the extent to which a measure is related specifically to theoretically targeted behaviors or experiences, and is not highly related to dissimilar behaviors or experiences, even if a similar measurement/assessment strategy (e.g., use of a questionnaire as a data collection strategy; Campbell & Fiske, 1959) is used in the different experiences. Campbell and Fiske (1959) posited multiple types of comparisons that provide evidence for the discriminant validity of measures, two of which are used here: the heterotrait–heteromethod and the monotrait–heteromethod approach. The first approach, the heterotrait–heteromethod approach, states that measures of the same trait using different methods should be more strongly related than measures of a different trait using a different method (for the purposes of this study, we call this *scale specificity*). A second, and more stringent, test of discriminant validity is the monotrait–heteromethod comparison that states that measures of the same trait using different methods should be more strongly related than measures of a different trait using the same method (for the purposes of this study, we refer to this approach as a *method variance contrast*). For example, inferences derived from a questionnaire measuring craving would be supported if the questionnaire items targeting craving were significantly related to real-time self-report measures of craving and less well related to questionnaire measures of a very different construct, such as depression proneness (see Figure 1).

In studies of nicotine dependence, construct validity is typically evaluated by correlating the new measure with the FTND (convergent validity; Heatherton et al., 1991) or with criteria from the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition, text revision (*DSM–IV–TR*; American Psychiatric Association, 2000) or by using the measure to predict smoking heaviness or relapse likelihood (predictive validity; e.g., Piper et al., 2004). Discriminant validity is rarely assessed.

Correlating a dependence questionnaire with dependence criteria such as smoking heaviness or relapse provides information on the clinical utility of the questionnaire. However, these relations do not provide much theoretical insight; that is, they do not reveal why the measure predicts relapse. They do not reveal whether the theory behind the measure might be correct. For instance, a dependence measure might elicit information on the severity of a person's withdrawal syndrome (e.g., the Drive scale of the NDSS; Shiffman et al., 2004). If this measure is then found to predict the tendency to relapse back to smoking, it would be tempting to speculate that the person relapses quickly because she or he experiences especially strong withdrawal symptoms if she or he abstains. However, this may not be the case. Broad dispositional factors or attitudes might, in fact, be responsible for the association of the measure with other dependence assays. For example, people who have low self-efficacy, who are high in neuroticism, or who believe that they are highly dependent might rate all sorts of dependence-related symptoms quite highly. Such people might also be more likely to relapse.

In the proposed research, we tested the ability of different questionnaire measures of nicotine dependence to assess relatively specific behavioral outcomes collected via real-time data acquisition. We selected this strategy for two reasons. First, evidence attests to the validity of real-time data acquisition methods; for example, they appear to index event occurrence more accurately than do retrospective recall methods (Stone et al., 1998; Stone, Shiffman, & DeVries, 1999). Therefore, such methods yield valuable validity information. Second, real-time assessment strategies entail different data acquisition methods than does the questionnaire method. As discussed above, the MTMM model of construct validation reveals that this sort of multimethod approach allows the investigator to examine associations among measures that are relatively uncontaminated by method variance (Campbell & Fiske, 1959; Cole, Martin, Peeke, Henderson, & Harwell, 1998). As compared with global questionnaires, ecological momentary assessment (EMA) usually involves assessment in real-world contexts, different item and response options, a different response method (e.g., use of a response slider bar), less temporal delay in reporting, less integration of data over time, and so forth. This means that comparisons of real-time and questionnaire methods should reduce the likelihood that agreement will be a function of reporting methods, broad attitudes, and recall or integration errors.

[1] Recently, there has been a debate in the measurement literature about whether constructs should be conceptualized via nomological networks. Boorsboom, Mellenbergh, and Van Heerden (2004) argued that measures gain validity to the extent that one can prove that (a) the construct exists and is worth measuring and that (b) changes in the construct cause changes in the measure of that construct. Thus, Boorsboom et al. viewed a construct as merely an independent variable that must be appraised in much the same way as any other sort of independent variable. However, even in this light, it would seem relevant to assess whether any independent variable or construct is more highly related to theoretically relevant variables than to other variables.

[2] Convergent and discriminant validity are only two of many tests that are used to assess the validity of a measure. New types of validity have been discussed recently, including, but not limited to, content, substantive, structural, generalizability, external, and consequential (Messick, 1995).

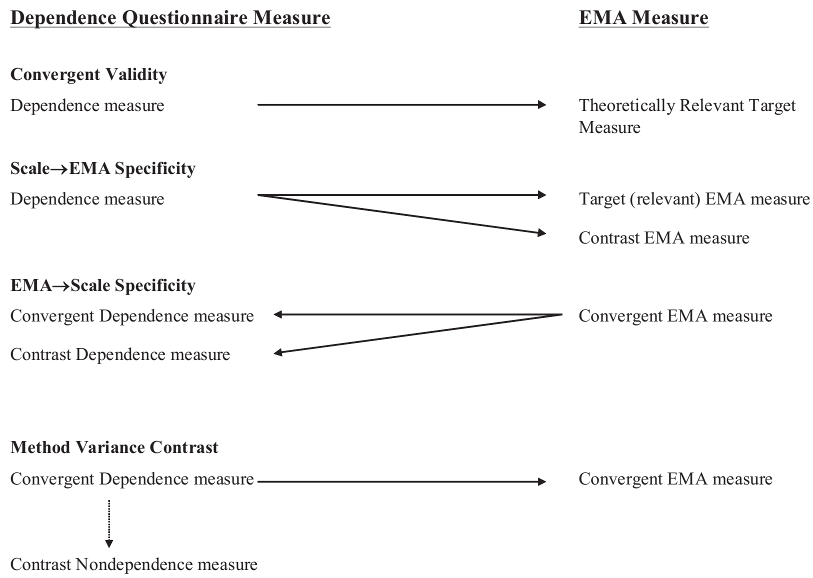

Figure 1. Overall analytical approach for construct validation of nicotine dependence scales. Convergent validity was tested by assessing the strength of the relation between a target nicotine dependence scale and a theoretically related ecological momentary assessment (EMA) measure. Discriminant validity was measured three ways: (a) scale-to-EMA specificity—comparing the strength of the relation between the nicotine dependence scale and a theoretically related EMA measure to the relation between the nicotine dependence scale and an unrelated EMA measure,(b) EMA-to-scale specificity—comparing the relation between a nicotine dependence scale to a theoretically related EMA measure with the relation between an unrelated nicotine dependence scale to the same EMA measure, (c) method contrast—comparing the relation between a nicotine dependence scale to a theoretically related EMA measure with the relation between the same nicotine dependence scale and an unrelated retrospective self-report measure.

As we have argued earlier, clinicians and theorists make decisions and inferences on the basis of the assumption that self-report measures correlate strongly with the particular behaviors the measures are designed to assess. It seems logical to assume that people can accurately report on their typical patterns of behavior. However, across many behavioral domains, self-report questionnaires have a mixed record: sometimes correlating poorly (e.g., Smith, Leffingwell, & Ptacek, 1999; Stone et al., 1998; Todd et al., 2005; Todd, Tennen, Carney, Armili, & Affleck, 2004) and sometimes well (pain: Jamison, Raymond, Slawsby, McHugo, & Baird, 2006; fatigue: Banthia et al., 2006) with real-time behavioral ratings. This variability may be because self-report measures ask people to report their global traits and to aggregate their behavioral tendencies over time, whereas in real life, people often behave in different ways across different situations (Mischel & Shoda, 1995). A more fine-grained, real-time/behavioral measure may capture more variability in behavior than does a questionnaire that targets global dispositions. Behavioral or real-time data, therefore, unlike global self-report questionnaires, may reflect complex Person × Situation interactions that constrain relations between a trait measure such as dependence and any behavior that varies across time and place (Bem & Allen, 1974; Bem & Funder, 1978; Mischel & Shoda, 1995). In the case of nicotine dependence, some studies have found that nicotine dependence measures are related to real-time reports of smoking heaviness or to cotinine levels (e.g., Chen et al., 2002; Pérez-Stable, Marín, Marín, Brody, & Benowitz, 1990; Prokhorov et al., 2000). Beyond this, little is known about whether dependence questionnaires accurately index particular tobacco use motives, symptoms, or behaviors as they occur in smokers' daily lives.

We evaluated the construct validity of three nicotine dependence measures in the present study by examining the relations of selected items and scales with symptoms and behaviors assessed via real-time data acquisition (i.e., EMA; Stone et al., 1998, 1999). The scales of the three questionnaires and the EMA items measured slightly different constructs using somewhat different assessment schemes. This research entailed determination of the construct validity of dependence questionnaire items via three sorts of relationships (see Figure 1): (a) the convergent validity of a dependence scale or item (i.e., whether a dependence measure predicts a theoretically related EMA symptom or behavior); (b) the scale specificity of a dependence scale or item or an EMA measure (i.e., a dependence scale's ability to predict a theoretically related EMA measure better than it predicts an unrelated EMA measure, or the lack of an association between a dependence scale and an unrelated EMA measure); and (c) the method variance contrasts of a dependence item or scale (i.e., the extent to which a self-report dependence questionnaire predicts a theoretically related EMA outcome better than it predicts an unrelated questionnaire measure). We hypothesized that the various dependence scales would show evidence of both convergent and discriminative validity as reflected in their relations with selected EMA and nondependence questionnaire measures but that the questionnaire scales and items with explicit behavioral referents (i.e., smoking heaviness) would show the strongest evidence of validity.

Method

Participants

Participants were 608 smokers (57.9% female, 76% Caucasian; see Table 1 for demographics) who were recruited as part of a clinical trial (Piper et al., 2007) conducted in Milwaukee, Wisconsin. Participants were recruited via flyers, and TV, radio, and newspaper advertisements. Inclusion criteria included smoking 10 or more cigarettes per day and being motivated to quit smoking. Exclusion criteria included any physical or psychological exclusion criteria, including Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977) scores greater than 16, heavy alcohol use, history of eating disorders, suicidality, or medical contraindications to bupropion or nicotine gum (e.g., history of seizure) that would prevent participating in the study. Female smokers could not be pregnant or breast-feeding and agreed to take steps to prevent pregnancy while taking study medication.

Procedure

Interested smokers called a central research office, where they completed a phone screen. Qualifying participants were invited to attend an orientation session where they provided written informed consent as well as demographic and smoking history information. Additional eligibility screening for medical and psychological contraindications was conducted at a baseline session. Participants also completed nicotine dependence self-report questionnaires at this visit, including the FTND (Heatherton et al., 1991), the NDSS (Shiffman et al., 2004), and the WISDM (Piper et al., 2004). Participants' responses to these questionnaires at baseline were used in all analyses.

Participants who passed the baseline screening were randomized to one of three treatment conditions: (a) active bupropion sustained release (SR) (150 mg, bid) + active 4-mg nicotine gum, (b) active bupropion SR + placebo nicotine gum, or (c) placebo bupropion + placebo gum. In addition to pharmacotherapy, participants received three brief (10-min) individual smoking cessation counseling sessions during which they set a quit date, were taught techniques to assist in maintaining abstinence, and were instructed in how to use the study medications. These three sessions took place 1 week prequit, on the quit day, and 1 week postquit. From 1 week prequit to 1 week postquit, participants carried cell phones that prompted them to respond to data probes in real time.

After the baseline session, there were six follow-up visits: one on the quit day and one each week for the first 3 weeks. After these first four visits, participants came into the clinic every other week for 2 weeks. Participants were also followed up via telephone at 6 and 12 months postquit.

Table 1
Demographics

Variable	Score
Gender (% female)	57.9
Race (% Caucasian)	76.0
Age: *M* (*SD*)	41.8 (11.3)
Cigarettes per day: *M* (*SD*)	22.4 (9.9)
Number of years smoked: *M* (*SD*)	24.0 (11.5)
FTND: *M* (*SD*)	5.6 (2.1)

Note. FTND = Fagerström Test for Nicotine Dependence.

Cell Phone Assessments

Participants responded to cell phone assessments during two scheduled prompts (morning and evening) and two random prompts throughout the day. The cell phone assessments were designed to collect data on smoking behavior, withdrawal, and affect in real time (EMA). EMA is fitting for this purpose because it is time stamped, specifying exactly when participants respond to questions. Previous studies have found that participants have been fairly reliable at responding to EMA prompts; 80%–88% of prompts were answered within 2–10 min (Csikszentmihalyi & Larson, 1987; Shiffman et al., 1997). In the present study, participants carried cell phones from 1 week prequit to 1 week postquit. The data were collected via an interactive voice recording system. When smoking was examined at the first call of the day, participants who reported smoking more than 10 cigarettes between the evening call and the morning call were excluded (thus, the top 5% of respondents were excluded from only these analyses) because it was felt that this degree of smoking during the night was unlikely, and therefore the participants had likely misunderstood the question.

Measures

The CES-D Scale. The CES-D Scale (Radloff, 1977) is a 20-item self-report questionnaire that assesses depressive symptoms. Possible scores on the scale range from 0 to 60, with higher numbers indicating greater depressive symptoms.

The FTND. The FTND (Heatherton et al., 1991) comprises six items. Scores on the scale range from 0 to 10, with higher numbers indicating greater dependence. The FTND has been validated across several studies and is generally found to have poor internal consistency (alphas below .80) and robust correlations with cigarettes per day, expired carbon monoxide, cotinine, and success in a smoking cessation attempt (e.g., Heatherton et al., 1991; Payne, Smith, McCracken, McSherry, & Anthony, 1994; Piper, McCarthy, & Baker, 2006).

The Michigan Alcoholism Screening Test. The Michigan Alcoholism Screening Test (MAST; Selzer, 1971) is a 24-item measure that can be administered as a self-report questionnaire (Skinner, 1979) to assess severity of problems with alcohol. Scoring followed the binary scoring procedure evaluated by Skinner (1979), with items scored either 0 (no) or 1 (yes).

The NDSS. The NDSS (Shiffman et al., 2004) is a 23-item self-report measure with five scales: Tolerance, Drive, Stereotypy, Continuity, and Priority. The NDSS was validated with a sample of 317 adult, heavy (>9 cigarettes/day for at least 2 years), daily, treatment-seeking smokers attending a research smoking cessation clinic. Internal consistency was good (α = 86 for the total scale). The NDSS total score and four of the NDSS subscales were correlated with the Fagerstrom Tolerance Questionnaire (a precursor to the FTND; Fagerström, 1978; NDSS total r = .54), cigarettes per day (r = .48), and cotinine levels (.27).

The WISDM. The WISDM (Piper et al., 2004) is a 68-item measure that assesses 13 theoretically derived motivational domains: Affiliative Attachment, Automaticity, Behavioral Choice/

Melioration, Cognitive Enhancement, Craving, Cue Exposure/Associative Processes, Loss of Control, Negative Reinforcement, Positive Reinforcement, Social and Environmental Goads, Taste and Sensory Properties, Tolerance, and Weight Control. The WISDM was validated using a sample of 775 participants from multiple sources, including participants in present studies, introductory psychology students, and community volunteers who were at least 18 years of age and who had smoked at least once in the past 2 weeks. All subscales had high internal consistence (αs >.90). Most scales were found to be significantly correlated with the FTND (*r*s = .01–.78). In addition, most scales were found to be significantly correlated with cigarettes smoked per day (*r*s = .23–.76), expired carbon monoxide levels (*r*s = .15–.74), and *DSM–IV* criteria for nicotine dependence (*r*s = .31–.73).

The Wisconsin Smoking Withdrawal Scale. The Wisconsin Smoking Withdrawal Scale (WSWS; Welsch et al., 1999) is a 28-item measure of nicotine withdrawal and comprises seven scales: Anger, Anxiety, Concentration, Craving, Hunger, Sadness, and Sleep. A condensed version of the WSWS was used in the EMA assessments that included 2 anger items, 2 anxiety items, 2 craving items, 2 hunger items, 2 sadness items, and 1 concentration item. These items were selected for the condensed version because they were the highest loading items on their individual factors in the factor analysis validating the scale (Welsch et al., 1999).

Data Analysis

Construct Validity Analyses

The present analyses do not constitute a full MTMM analysis (Campbell & Fiske, 1959), in part because our goal was to validate only the nicotine dependence measures. We did not intend to evaluate the validity of the measures administered via EMA; thus, the evaluations for that side of the matrix.were not completed. We also did not have two identical sets of trait-bymethod comparisons; thus different questionnaires were used for the method variance contrasts and scale specificity analyses and for the convergent validity comparisons. Another caveat is that the different questionnaire-EMA facets not only reflect the molar differences in targeted trait and general assessment strategy (i.e., questionnaire vs. EMA) but also reflect differences in fine-grained features such as time frame and response scale. Because of these complexities, the focus was on analytic comparisons that address the validity of the questionnaire nicotine dependence measures, the MTMM model was used (Campbell & Fiske, 1959) as an informal guiding strategy.

Convergent Validity

Motives measures. Nicotine dependence questionnaire items or scales were used to predict EMA items. One challenge of this study was identifying EMA items that were especially relevant to the dependence measures to be tested. The construct validity of only those global self-report scales was tested for which theoretically related EMA items were available. Intrinsic to the nature of many dependence measures (e.g., smoking for negative reinforcement or cognitive enhancement) is the notion that the dependent individual smokes to control, reduce, or ameliorate a particular symptom or phenomenon; that is, these measures assess *motives* for smoking. Therefore, we posited that if individuals who reported a specific motive for smoking (e.g., high scores on WISDM Cognitive Enhancement, Craving, Weight Control, Negative Reinforcement or NDSS Drive) refrained from smoking, they would experience a spike in the relevant symptoms due to the loss (or perceived loss) of this regulatory or control strategy. For example, if smokers indicated they smoked to control negative affect, then we tested whether they showed a strong increase in negative affect in response to abstinence. This was tested via correlations between global self-report dependence scales and postquit EMA scores from which prequit EMA scores had been statistically partialled.[3,4] It is important to note that this strategy is conservative because it does not merely ask the respondent to endorse beliefs about smoking in real time that match the beliefs elicited by the dependence questionnaire. Rather, it addresses whether questionnaire data gathered on specific dependence motives predict an outcome of clinical importance (severity of withdrawal, smoking rates, and patterns) that is a downstream consequence of the causal processes addressed in the dependence measure.

Cigarette use measures. Some measures elicited information about particular behaviors: for example, FTND Item 1 (time to first cigarette in the morning), FTND Item 4 (cigarettes per day), and the WISDM and NDSS Tolerance scales (that elicit information about smoking heaviness). These items/scales were used to predict theoretically related items from the EMA data prequit, before participants began altering these behaviors as a consequence of the quit attempt (e.g., number of cigarettes smoked per day [FTND Item 4] was used to predict real-time measures of precessation smoking heaviness in the week after participants completed the FTND questionnaire).[5]

Control variables. Linear regression was used for all analyses, and gender and treatment condition were entered as covariates. Treatment was coded with two dummy variables that coded for mono and combination pharmacotherapy, with the placebo control

[3] We also conducted all analyses with gain scores computed by subtracting prequit from postquit scores on the EMA measure. These analyses yielded a very similar pattern of findings to that obtained with residualized scores. Because the relation between a variable and a measure of change is highly influenced by such factors as the correlation between the pre- and postscores constituting the change variable (e.g., Gardner & Neufeld, 1987), we not only examined such relations but also demonstrated the stability of solutions using both gain and residualized scores.

[4] An alternative strategy would have been to relate global dependence scales to EMA-assessed changes in response to smoking individual cigarettes (e.g., the extent to which negative affect changes from a precigarette to a postcigarette). However, the temporal resolution of EMA reports was not ideal for this sort of analysis. This would have required pre- and postratings of individual cigarettes, and because EMA data were collected only four times a day, we were unable to conduct such analyses. We also did not want to relate EMA symptom ratings with smoking reports that were gathered at the same measurement occasion (relating symptom reports with whether a person had just smoked) because such simultaneous ratings raise concerns about reciprocal causality.

[5] In this instance, we were examining the agreement of two measures of a particular behavior (cigarettes smoked per day). These analyses could be viewed merely as an instance of predictive validity as opposed to construct validity; however, one could also view the smoking measures as reflecting a larger construct of smoking heaviness that, in theory, could be reflected by biochemical, topography, and self-report measures. The chief distinction between the cigarette use predictions and the other analyses, therefore, is the availability of an objective marker (cigarette count).

group coded as the reference condition. Analyses of dependence motives also controlled for level of dependence (measured via cigarettes per day)[6]; this was done to partial variance due to general versus specific dependence motives (such as smoking specifically for cognitive enhancement). For the WISDM Weight Control analyses only, analyses also controlled for the number of days the participant reported smoking one or more cigarettes from the quit day to the time point at which weight gain was assessed. This was done because of evidence that smoking suppresses both hunger between meals and weight and could therefore bias the relation between the Weight Control scale and the hunger and weight gain measures (Hudmon et al., 1999; Klein & Corwin, 2004; Klesges et al., 1997). For other tests of specific behavioral manifestations of dependence, the cigarettes per day variable was not used as a covariate because this variable was thought to be conceptually central to the other behavioral EMA criteria (e.g., latency to smoke in the morning appears to reflect the same latent variable as cigarettes per day; see smoking heaviness index in Baker et al., 2007; Heatherton et al., 1991; Lessov et al., 2004).

Time coding. The *prequit period* was defined a priori as 7–4 days prequit, and the mean of all EMA responses for each item or scale given on those days was calculated, unless a measure specifically referred to smokers' experiences at a specific time of day (e.g., FTND Item 1). The 3 days immediately preceding the quit day were excluded from analyses because prior research showed changes in symptom trajectories within a day or two prior to the quit day (e.g., McCarthy, Piasecki, Fiore, & Baker, 2006), suggesting that such data would not be representative. The *postquit period* was defined a priori as days 1–3 postquit, and the mean of all EMA responses for each given item or scale for those days was calculated. EMA reports from 4 to 7 days postquit were not included because withdrawal and symptomatic rebound were expected to be most intense during the first few days of the quit attempt (Shiffman et al., 2006; Shiffman & Sayette, 2005), and later time points were likely to be more affected by postquit smoking as participants relapsed.[7] Also, previous data suggest that early withdrawal reactions (within the first several days) reflect the largest withdrawal effects (e.g., McCarthy et al., 2006). The prequit assessments resulted in 5,536 EMA reports; the postquit assessments resulted in 5,077 reports.

Procedure. The order of entry of predictors in the regression models was as follows: control variables, the relevant EMA prescore, and then the dependence scale/item of interest. Convergent validity was conceptualized as the increment in variance accounted for in the residualized dependent variable by the dependence variable, over and above the variance accounted for by the control variables (the change in the r^2 value of the model when the dependence scale was added; Δr^2).

Discriminant Validity

Two types of discriminant validity were examined in this study: scale specificity and method variance contrasts. Discriminant validity was assessed only for scales that showed significant relationships with theoretically related EMA measures.

Scale specificity. To test for scale specificity discriminant validity, EMA measures were selected so that one measure was highly theoretically relevant to the dependence questionnaire measure, whereas the *contrast* EMA measure was highly dissimilar. For example, for the dependence motives scales, a WSWS scale was chosen (i.e., WSWS hunger assessed via EMA) as the contrast scale. In theory, a dependence questionnaire measure of smoking to relieve negative affect (e.g., the WISDM Negative Reinforcement scale) should be more highly related to EMA-reported negative affect than it would be related to EMA-reported hunger (as suggested by prior research examining relations between affect and hunger measures among smokers attempting to quit smoking; e.g., Welsch et al., 1999). The smoking behavior EMA measures tended to be count measures (e.g., cigarettes per day). Thus, the only nonsmoking-related categorical measure was chosen in the EMA reports to test for scale specificity: "Had a stressful event since last call?" Evidence of scale specificity was inferred to the extent that the relation between the nicotine dependence measure and the theoretically related EMA measure was stronger than the relation between the same nicotine dependence scale and a theoretically unrelated EMA measure (i.e., hunger or stressor occurrence). Both with regards to the hunger and stressor EMA measures, item interrcorrelations showed low levels of correlation with the EMA measures thought to reflect dependence motives (e.g., craving, negative affect).

In addition to evaluating whether the nicotine dependence measure was related to a theoretically unrelated EMA measure, an assessment was made of whether a theoretically unrelated nicotine dependence scale was related to the target EMA measure. It is possible that the target EMA measure reflected a general feature of dependence and would be similarly predictable by any dependence questionnaire. To evaluate this, it was assessed whether a theoretically unrelated dependence questionnaire measure was as predictive of the EMA measure used in the convergent validity analyses as was a theoretically relevant questionnaire measure. Only three comparison nicotine dependence questionnaire scales were chosen for the convergent validity analyses. Many measures of nicotine dependence are highly related to one another on a substantive basis; most measures of nicotine dependence should be highly associated with one another in that they measure the same global construct. However, a decision was made to use nicotine dependence measures because their response scales were identical to the target nicotine dependence scales, and it was possible to identify three that were believed to be conceptually distinct from the targeted dependence measures: the WISDM Taste/Sensory Properties scale (smoking due to a preference for the taste of cigarettes), the WISDM Social/Environmental Goads scale (having a lot of friends who smoke), and the NDSS Continuity scale (smoking the same amount each day). EMA scale specificity was inferred if an EMA measure was more highly related to a conceptually related global dependence scale/item than to a conceptually

[6] When analyses of dependence motives were conducted without controlling for cigarettes per day, the pattern of significant effects remained the same.

[7] Because we collapsed data over time, the present analyses do not permit a test of the influence of situations on the relation between dependence measures and behavioral outcomes. It may be that if situations were treated formally, and not as error, then higher correlations would be obtained between dependence questionnaires/items and EMA ratings (e.g., Mischel & Peake, 1983). However, for many of the measures, the nature of the data did not permit separate analysis of questionnaire–EMA relations as a function of a situational typology.

unrelated comparison dependence questionnaire scale/item. EMA scale specificity was assessed only for scales that showed significant relations with theoretically related EMA measures (i.e., possessed convergent validity).

Method variance contrasts. The final test of discriminant validity was to determine whether the shared variance between different types of measures of the same trait exceeded the amount of shared variance between measures of different traits using similar methods. For these analyses, two theoretically unrelated comparison questionnaire measures were used: the MAST and CES-D. These measures were selected because—like the targeted nicotine dependence questionnaire measures—they are retrospective self-report measures. The two measures used somewhat different response scales: the MAST uses a yes–no checklist, whereas the CES-D uses a Likert-type response scale. For these comparisons, the strength of the relation between the targeted nicotine dependence item/scale and its theoretically related EMA scale was compared with the size of the relation between the same nicotine dependence item/scale and the comparison retrospective questionnaire measures (the MAST or CES-D).

Results

Convergent Validity

Tests of motives scales. The motives scales tended to predict pre–post increases in theoretically related EMA items and to account for variance over and above that accounted for by control variables (see Table 2).[8] In this regard, the WISDM Cognitive Enhancement scale (e.g., "I smoke when I really need to concentrate") predicted an increase in difficulty concentrating experienced by participants postquit. Similarly, the WISDM Negative Reinforcement scale (e.g., "Smoking a cigarette improves my mood") predicted emergent negative affect postquit, and the WISDM and NDSS craving scales predicted emergent craving postquit.[9] The WISDM Weight Control scale, however, predicted neither postquit hunger in the first several days postquit nor weight gain at 4 or 8 weeks postquit. Even though relations between predictors and theoretically related EMA outcomes tended to be significant, the amount of variance accounted for by the scales was modest (e.g., accounting for only a 2%–4% reduction in error: see Table 2).[10]

Tests of smoking behavior measures. Next, we examined the convergent validity of dependence measures that targeted smoking behaviors. As with the dependence motives scales, there were significant relations between the dependence measures and theoretically related EMA items; the variance accounted for by such variables tended to be higher than with the motives measures (reductions in the residual ranged from 2% to 37%; see Table 2). FTND Item 1 ("How soon after waking do you smoke your first cigarette?") predicted the EMA prequit morning report of the number of cigarettes smoked since the last call (which had been completed immediately before bed). FTND Item 4 ("How many cigarettes do you smoke per day?") predicted EMA prequit reports of number of cigarettes smoked per day. Both the NDSS and WISDM Tolerance scales also predicted the number of cigarettes smoked per day during the prequit period (see Table 2).

Accuracy of high-risk group selection. To explore the amount of information that could be obtained by such modest but statistically significant relations, we considered whether the present level of convergent validity was sufficient to select accurately individuals in the top quartile of symptomatic or behavioral severity. Thus, we determined the percentage of people who scored in the top 25% on a dependence assessment item or scale and who also scored in the top 25% in terms of their level on the theoretically related EMA measure. First, we computed this analysis for the variable pair with the smallest, significant relation with one another (see Table 2): WISDM Negative Reinforcement and WSWS Negative Affect. Of those who scored in the top 25% on WISDM Negative Reinforcement, 37.4% scored within the top 25% on WSWS postquit Negative Affect. Next, we computed this analysis for a variable pair with a moderate level of association: global NDSS Drive and EMA-assessed WSWS Craving. Of those who scored in the top 25% on the NDSS Drive scale, 40.6% scored in the top 25% on the WSWS Craving scale. Finally, we analyzed the pair of scores with the highest level of association: FTND cigarettes per day and EMA report of number of cigarettes smoked prequit (see Table 2). Of those who scored in the highest 25% for number of cigarettes per day on the FTND, 69.1% scored in the highest 25% on the EMA report of number of cigarettes smoked. Thus, these sensitivity-type analyses suggested that only some of the targeted relations are sufficiently strong so as to afford clinically useful predictions.

Scale Specificity Discriminant Validity

Dependence measures. Our next goal was to assess the scale specificity of the nicotine dependence questionnaire measures. As expected, most of the motives scales (WISDM Cognitive Enhancement, WISDM Negative Reinforcement, and NDSS Drive) failed

[8] We also examined whether the dependence motives measures would predict theoretically related symptoms prequit and postquit, in addition to predicting the increase from pre- to postquit. All of the motives measures predicted theoretically related prequit and postquit symptoms. However, because the dependence motives measures assess whether smoking helps to control certain symptoms, we felt that the postquit increase in such symptoms was a more relevant and demanding test.

[9] This was one instance in which the subtractive gain score analysis diverged from the residualized score analysis in that relations between the craving dependence scales and emergent craving ratings were weaker and/or borderline significant with the gain scores.

[10] To examine the upper limit of these types of comparisons, we compared WSWS ratings during the first 3 days postquit with retrospective WSWS ratings at the Week 1 visit (at approximately 7 days postquit) where participants were asked to rate how they felt over the past week. To make these comparisons as strong as possible, we used only the WSWS items administered in the palm pilot assessments and assessed the amount of variance accounted for in the models over and above the control variables of gender, treatment condition, and cigarettes per day. We found that the WSWS craving measure collected at Week 1 accounted for 35.6% of the variance in real-time measures of craving over the first 3 days postquit; WSWS negative affect ratings at Week 1 accounted for 25.5% of the variance in real-time reports of WSWS negative affect; and WSWS hunger at Week 1 accounted for 35.6% of the variance in real-time reports of WSWS hunger. These findings suggest that a single global measure could predict meaningful amounts of variance in the phasic EMA measures; that is, variation in phasic EMA measures over time did not preclude meaningful associations.

Table 2
Convergent Validity of Nicotine Dependence Measures With Theoretically Related Outcomes (Using Linear Regression, Controlling for Gender and Treatment)

Dependence measure	EMA measure	Time point	Δr^{2a}	Percent decrease in residual	β	*t*	*p*
WISDM Cognitive Enhancement	WSWS Concentration	Postquit	.02	2.81	.15	3.86	<.001
WISDM Craving	WSWS Craving	Postquit	.03	3.53	.18	4.34	<.001
NDSS Drive	WSWS Craving	Postquit	.03	4.01	.09	1.97	<.001
WISDM Negative Reinforcement	WSWS Negative Affect	Postquit	.01	2.10	.09	3.33	.001
WISDM Weight Control	WSWS Hunger	Postquit	.003	0.41	.06	1.47	.14
	Change in weight since baseline	Postquit 4 weeks	.001	0.001	.03	0.65	.51
	Change in weight since baseline	Postquit 8 weeks	.002	0.25	−.05	−0.96	.34
FTND Item 1	Number of cigarettes since last call	Prequit morning	.03	2.67	.16	3.68	<.001
	Cigarettes since last call	Postquit morning	.02	2.05	.14	3.24	.001
FTND Item 4	Cigarettes since yesterday	Prequit	.36	37.17	.61	17.61	<.001
NDSS Tolerance	Cigarettes since yesterday	Prequit	.10	10.68	.32	7.89	<.001
WISDM Tolerance	Cigarettes since yesterday	Prequit	.09	9.53	.31	7.43	<.001

Note. EMA = ecological momentary assessment; WISDM = Wisconsin Inventory of Smoking Dependence Motives; WSWS = Wisconsin Smoking Withdrawal Scale; FTND = Fagerström Test of Nicotine Dependence; NDSS = Nicotine Dependence Syndrome Scale.
[a] Δr^2 denotes the change in r^2 between the model with only control variables and the model with the nicotine dependence measure added.

to predict the EMA measure of postquit increases in hunger (see Table 3). The single exception to this is that the WISDM Craving scale predicted residualized postquit hunger.

A more stringent test of discriminant validity concerns whether the association between the dependence scale scores are significantly more highly associated with the theoretically linked EMA scores than they are with the theoretically unrelated EMA scores. To test this, we used the procedure proposed by Meng and colleagues (Meng, Rosenthal, & Rubin, 1992) for testing differences between correlations using the *z* normal curve test. This procedure showed that WISDM Cognitive Enhancement was not significantly more strongly related to WSWS Concentration than it was to WSWS Hunger. Similarly, WISDM Negative Reinforcement was not significantly more strongly related to WSWS Negative Affect than it was to WSWS Hunger. However, both the NDSS Drive scale and the WISDM Craving scale were more strongly related to WSWS Craving than they were to WSWS Hunger.

As predicted, none of the cigarette use scales (WISDM and NDSS Tolerance scales, FTND Items 1 or 4) predicted stressful events since last call. Therefore, although these dependence mea-

Table 3
Scale Specificity Discriminative Validity for the Dependence Measures

Dependence measure	EMA measure	Time point	Δr^{2a}	Percent decrease in residual	β	*t*	*p*
WISDM Cognitive Enhancement	WSWS Concentration	Postquit[b]	.02	2.81	.15	3.86	<.001
	WSWS Hunger	Postquit	.004	0.62	.06	1.80	.07
WISDM Negative Reinforcement	WSWS Negative Affect	Postquit	.01	2.10	.09	3.33	.001
	WSWS Hunger	Postquit	.002	0.21	.05	1.04	.30
NDSS Drive	WSWS Craving	Postquit	.02	2.81	.15	3.86	<.001
	WSWS Hunger[c]	Postquit	.004	0.62	.06	1.80	.07
WISDM Craving	WSWS Craving	Postquit	.03	3.53	.18	4.34	<.001
	WSWS Hunger[c]	Postquit	.007	2.12	.09	2.43	.02
WISDM Tolerance	Cigarettes since yesterday	Prequit	.09	9.53	.31	7.43	<.001
	Stressful situation since last call[c]	Prequit	.003	0.32	−.06	−1.31	.19
NDSS Tolerance	Cigarettes since yesterday	Prequit	.10	10.68	.32	7.89	<.001
	Stressful situation since last call[c]	Prequit	.000	0.00	.003	.06	.95
FTND Item 1	Cigarettes since last call	Prequit morning	.02	2.05	.14	3.24	.001
	Stressful situation since last call[c]	Prequit	.00	0.01	.01	.22	.83
FTND Item 4	Cigarettes since yesterday	Prequit	.36	37.17	.61	17.61	<.001
	Stressful situation since last call[c]	Prequit	.002	0.25	−.05	−1.14	.26

Note. EMA = ecological momentary assessment; WISDM = Wisconsin Inventory of Smoking Dependence Motives; WSWS = Wisconsin Smoking Withdrawal Scale; NDSS = Nicotine Dependence Syndrome Scale; FTND = Fagerström Test of Nicotine Dependence.
[a] Δr^2 denotes the change in r^2 between the model with only control variables and the model with the nicotine dependence measure added. [b] The tests using dependent variables measured postquit were conducted controlling for prequit levels of the dependent variable. [c] The theoretically related scale is significantly more strongly related to the dependent variable than the comparison scale.

sures predicted EMA cigarette use items from theoretically relevant domains, they did not predict EMA items from conceptually distinct domains. The *z* test for comparison of the correlations showed that the dependence measures were all significantly more highly related to the smoking behavior measures than to the stressful events measure ($ps < .01$).

EMA measures. Our next goal was to assess the scale specificity discriminant validity of the EMA target measures (see Table 4). In general, the analyzed measures yielded positive evidence of discriminant validity. WISDM Cognitive Enhancement, WISDM Negative Reinforcement, WISDM Craving, and the NDSS Drive scales had stronger relations with their relevant EMA measures than did theoretically unrelated comparison questionnaire measures. Table 5 shows that the theoretically relevant dependence measures produced percentage reductions in the residuals that were often 10 times the magnitude of those produced by the unrelated dependence scales. The *z* test for differences among the correlations revealed that the theoretically linked associations were significantly stronger than the comparison associations except in three instances. The relations between the WISDM Craving and Taste scales did not differ significantly in terms of their associations with the EMA Craving increase ($p > .05$). Similarly, of the three possible comparisons of correlation magnitude for the EMA negative affect measure, the relations of EMA negative affect with the WISDM Negative Reinforcement scale significantly exceeded only its relations with the WISDM Social and Environmental Goads scale (and not its associations with the WISDM Taste or the NDSS Continuity scales).

There was substantial evidence of discriminant validity with regards to the cigarette use measures. For instance, the EMA measure of cigarettes smoked per day was highly related to FTND Item 4 (cigarettes per day) and to the two tolerance scales from the NDSS and the WISDM. The EMA measure of cigarettes per day was also significantly related to the WISDM Taste/Sensory Properties scale and the NDSS Continuity scale, but the relations were much weaker. Morning smoking was predicted only by FTND Item 1 (time to first cigarette in the morning), consistent with the behavior targeted by that item. These characterizations were consistent with the results of *z* tests for comparison of the correlations: The theoretically congruent associations were consistently higher than the comparison associations ($ps < .05$).

Method Contrasts Discriminant Validity

The motives scales showed mixed support for the method contrasts discriminant validity. Although all of the dependence measures were significantly related to theoretically linked EMA measures (see Table 5), the magnitudes of these associations did not always exceed those between the dependence scales and the comparison questionnaire measures. The *z* tests of correlation magnitude revealed that WISDM Cognitive Enhancement was more strongly related to EMA reports of increased difficulty concentrating postquit than it was to a questionnaire measure of alcohol dependence (the MAST). However, WISDM Cognitive Enhancement was about equally as strongly related to retrospective self-reports of depression symptoms (the CES-D), a questionnaire with a more similar response

Table 4
Scale Specificity Discriminant Validity for the Ecological Momentary Assessment (EMA) Measures

EMA dependent variable	Dependence measure (independent variable)	Δr^{2a}	Percent decrease in residual	B	*t*	*p*
WSWS Concentration (postquit[b])	**WISDM Cognitive Enhancement**	**.02**	**2.81**	**.15**	**3.86**	**<.001**
	WISDM Social/Environmental Goads[c]	.00	0.01	−.01	−0.24	.81
	WISDM Taste[c]	.002	0.25	.04	1.13	.26
	NDSS Continuity[c]	.001	0.05	.02	0.51	.61
WSWS Craving (postquit)	**NDSS Drive**	**.03**	**4.01**	**.09**	**1.97**	**<.001**
	WISDM Craving	**.03**	**3.53**	**.18**	**4.34**	**<.001**
	WISDM Social/Environmental Goads[c]	.002	0.23	−.04	−1.09	.28
	WISDM Taste[c]	.006	0.77	.08	2.00	.05
	NDSS Continuity[c]	.00	0.03	.01	0.36	.72
WSWS Negative Affect (postquit)	**WISDM Negative Reinforcement**	**.01**	**2.10**	**.09**	**3.33**	**.001**
	WISDM Social/Environmental Goads[c]	.001	0.27	−.03	−1.27	.24
	WISDM Taste	.002	0.73	.05	1.95	.05
	NDSS Continuity	.004	1.13	.06	1.27	.21
Cigarettes per day	**WISDM Tolerance**	**.09**	**9.53**	**.31**	**7.43**	**<.001**
	NDSS Tolerance	**.10**	**10.68**	**.32**	**7.89**	**<.001**
	FTND Item 4	**.36**	**37.17**	**.61**	**17.61**	**<.001**
	WISDM Social/Environmental Goads[c]	.006	0.62	.08	1.80	.07
	WISDM Taste[c]	.02	1.52	.12	2.85	.005
	NDSS Continuity[c]	.02	1.92	.14	3.19	.002
Cigarettes since last call (morning call prequit)	**FTND Item 1**	**.02**	**2.05**	**.14**	**3.24**	**.001**
	WISDM Social/Environmental Goads[c]	.001	0.02	.02	0.33	.75
	WISDM Taste[c]	.002	0.18	.04	0.95	.34
	NDSS Continuity[c]	.002	0.19	.04	0.96	.34

Note. WSWS = Wisconsin Smoking Withdrawal Scale; WISDM = Wisconsin Inventory of Smoking Dependence Motives; NDSS = Nicotine Dependence Syndrome Scale; FTND = Fagerström Test of Nicotine Dependence. Boldface type indicates the relationship hypothesized to be the strongest if the scale has construct validity.
[a] Δr^2 denotes the change in r^2 between the model with only control variables and the model with the nicotine dependence measure added. [b] The tests using dependent variables measured postquit were conducted controlling for prequit levels of the dependent variable. [c] The theoretically related scale is significantly more strongly related to the dependent variable than the comparison scale.

Table 5
Method Variance Contrast Discriminant Validity

Dependence measure	Comparison measure	Δr^{2a}	Percent decrease in residual	β	*t*	*p*
WISDM Cognitive Enhancement	**EMA WSWS Concentration postquit**[b]	**.02**	**2.81**	**.15**	3.86	<.001
	MAST[c]	.00	0.00	−.01	−.13	.90
	CES-D	.03	2.59	.16	3.71	<.001
NDSS Drive	**EMA WSWS Craving postquit**	**.03**	**4.01**	**.09**	**1.97**	**<.001**
	MAST[c]	.004	0.39	.06	1.41	.16
	CES-D[c]	.002	0.23	.05	1.09	.28
WISDM Craving	**EMA WSWS Craving postquit**	**.03**	**3.53**	**.18**	**4.34**	**<.001**
	MAST[c]	.001	0.00	−.01	−0.27	.78
	CES-D[c]	.004	0.42	.07	1.48	.14
WISDM Negative Reinforcement	**EMA WSWS Negative Affect post–pre**	**.01**	**2.10**	**.09**	**3.33**	**.001**
	MAST	.006	0.61	.08	1.78	.08
	CES-D	.04	3.62	.19	4.41	<.001
FTND Item 1	**Number of cigarettes since last call prequit morning**	**.02**	**2.05**	**.14**	**3.24**	**.001**
	MAST	.002	0.21	.05	1.02	.31
	CES-D[c]	.002	0.16	.04	0.88	.38
FTND Item 4	**Cigarettes since yesterday prequit**	**.36**	**37.17**	**.61**	**17.61**	**<.001**
	MAST[c]	.00	0.00	.00	−0.05	.96
	CES-D[c]	.00	0.00	.02	0.43	.67
NDSS Tolerance	**Cigarettes since yesterday prequit**	**.10**	**10.68**	**.32**	**7.89**	**<.001**
	MAST[c]	.00	0.01	.03	0.64	.52
	CES-D[c]	.001	0.01	−.03	−0.64	.52
WISDM Tolerance	**Cigarettes since yesterday prequit**	**.09**	**9.53**	**.31**	**7.43**	**<.001**
	MAST[c]	.001	0.17	.04	0.95	.34
	CES-D[c]	.005	0.53	.07	1.69	.10

Note. WISDM = Wisconsin Inventory of Smoking Dependence Motives; EMA = ecological momentary assessment; WSWS = Wisconsin Smoking Withdrawal Scale; MAST = Michigan Alcoholism Screening Test; CES-D = Center for Epidemiologic Studies Depression scale; FTND = Fagerström Test of Nicotine Dependence; NDSS = Nicotine Dependence Syndrome Scale. Boldface type indicates the scale hypothesized to have the strongest relationship with the dependence measure because of their shared constructs.
[a] Δr^2 denotes the change in r^2 between the model with only control variables and the model with the nicotine dependence measure added. [b] The tests using dependent variables measured postquit were conducted controlling for prequit levels of the dependent variable. [c] The theoretically related scale is significantly more strongly related to the dependent variable than the comparison scale.

scale than the alcohol dependence measure. Both the NDSS Drive and the WISDM Craving scales were more strongly related to postquit EMA reports of craving than to questionnaire measures of alcohol dependence or depression (*p*s <.05). The WISDM Negative Reinforcement scale was more strongly related to postquit increases in negative affect than it was to the questionnaire measures of alcohol dependence, but less strongly related than it was to the questionnaire measure of depression symptoms. However, depression is clearly related to smoking for negative reinforcement (e.g., to relieve negative affect), and thus is probably a nonoptimal comparison scale. The cigarette use scales faired very well in discriminant validity. The FTND cigarette use items and the WISDM and NDSS Tolerance scales were more strongly related to real-time measures of cigarette use than they were to questionnaire measures of alcohol dependence and depression (*p*s < 05).

Discussion

The present study had two main goals: (a) to determine whether nicotine dependence self-report questionnaires reflect what smokers actually do and experience in their daily lives as evaluated via EMA measures and (b) to determine whether such nicotine dependence measures possess discriminant validity.

To address the first goal, we used scales and individual items from three nicotine dependence questionnaires to predict relevant, real-time, EMA self-reports of symptoms and behaviors. Overall, the nicotine dependence measures predicted theoretically related real-time EMA outcomes. However, the percentage of variance accounted for in the relevant behaviors and symptoms was fairly modest: For example, the relevant dependence questionnaire measures accounted for about 2%–3% of variance in EMA-measured emergent withdrawal symptoms. The dependence measures targeting smoking behavior performed considerably better (accounting for 10%–37% of the variance). The level of association between the nicotine dependence measures and the EMA items provides some support for the theoretical basis of the dependence scales and items. However, the magnitude of associations between these two types of variables is sufficiently weak in some cases (e.g., for negative reinforcement motives) that the scales have limited clinical utility for identifying which smokers will show extreme changes on relevant behaviors or symptoms during a quit attempt.

There are many plausible explanations for the modest predictive relations between the dependence scales/items and the EMA reports. First, different response scales and different time frames for responding are used in the two types of measures. Thus, we would expect some differences in method variance to constrain their association. Second, it may not be primarily experience with tobacco use that influences participants' responses to nicotine dependence questionnaires. For instance, participants' nicotine dependence questionnaire responses may reflect their mistaken attributions that tobacco abstinence causes their affective or cognitive symptoms, when in fact these symptoms may reflect other causes such as traitlike characteristics that are independent of nicotine dependence (e.g., neuroticism) and that do not affect EMA ratings similarly. There is a wealth of evidence from health psychology that such symptom attributions occur and can affect questionnaire and diagnostic assessments (e.g., Leventhal, Weinman, Leventhal, & Phillips, 2008). Third, the EMA items were somewhat conceptually distant from the dependence items/scales in that the dependence scales did not explicitly ask about how much symptoms or behaviors would escalate upon cessation, whereas the EMA items did reflect such pre- and postcessation effects. Fourth, the EMA items elicited responses about particular episodes, whereas the dependence scales/items elicited global, traitlike ratings—it may be that relations would have been higher had the EMA assessments synthesized data over many quit attempts. As discussed earlier, actual behavior may not be accurately reflected in global, traitlike ratings because people may behave differently in different situations (Mischel & Shoda, 1995). In addition, the very transient nature of the targeted symptoms means that asking people to provide global, traitlike ratings on a questionnaire or even averaging across episodes may not appropriately model the relations between individuals' perceptions of the frequency of these symptoms and the actual report of these symptoms (Heiby, 1995). Other analytical methods, including nonaveraged scores and nonlinear dynamic mathematical modeling, might have resulted in stronger relations. It is also possible that people differ in their ability to predict their behaviors or symptoms, such that more impressive associations would be found in a subgroup of people (Bem & Allen, 1974; Bem & Funder, 1978; Mischel & Shoda, 1995). Finally, it may be that some dependence measures are intrinsically weakly related to abstinence-induced change. This weak relationship may be because smokers have little idea of the extent to which smoking actually controls their symptoms, such as negative affect or craving.

Researchers have reported modest relations between global self-report measures and specific real-time measures in diverse research areas (e.g., coping; Smith et al., 1999; Stone et al., 1998; Todd et al., 2004; and drinking to cope; Todd et al., 2005). Some studies, however, have shown high correlations between retrospective self-report measures and daily assessments (Banthia et al., 2006; Jamison et al., 2006). These studies assessed retrospective self-report compared with daily reports during the same period. The present study also showed impressive relations between global, retrospective self-report and an amalgam of the phasic EMA reports (see Footnote 10 where retrospective recall at 1-week postquit was meaningfully related to the phasic EMA symptoms gathered over 1–3 days postquit). These high correlations suggest there is enough reliable variance in the phasic EMA measures to permit their accurate prediction. However, either the participants in this research had little idea what their withdrawal would be like or some dependence questions elicited responses largely irrelevant to the effects of withdrawal.

An additional finding of note is that the dependence measures targeting tobacco use account for greater variance in the EMA outcomes than do the motives scales. This difference could be because tobacco use questionnaires tap a more easily quantifiable (or objective) variable. It may also be that it is clearer to participants exactly what they are being asked to report (e.g., cigarettes per day). When asking participants to count their cigarette use, there is likely less room for interpretation on the part of the participant, whereas Likert scales are subject to response style biases (Bolt & Johnson, in press). It is interesting to note that the two studies that found strong relations between global measures and daily diary measures assessed specific symptoms (fatigue: Banthia et al., 2006; pain: Jamison et al., 2006), whereas the studies that found less clear relationships assessed coping (e.g., coping: Smith et al., 1999; Stone et al., 1998; Todd et al., 2004; and drinking to cope: Todd et al., 2005). Such questions may be less concrete and therefore demand that the participant make an attribution, leading to greater difficulty providing an accurate report. It is also possible that the dependence measures targeting tobacco use showed more impressive associations because they were tested with EMA variables that were gathered precessation; that is, they did not entail extrapolation from prequit measures to postquit events.

Although the associations between some dependence questionnaire scales and EMA symptom ratings were modest, some of the dependence scales and items were significantly related to theoretically relevant symptoms and behaviors recorded in real time. These associations were significantly stronger in most cases than were those between the dependence scales and theoretically unrelated symptom or behavioral domains. These findings provide some support for theoretical inferences made on the basis of some of the scales (e.g., negative reinforcement). However, the level of the associations suggests that the dependence scales that tapped individual motives were too inaccurate to support the clinical use of those scales for the purpose of postquit prognostication. This finding raises questions about the usefulness of some of the information counselors extract in counseling sessions with smokers because smokers' characterization of their smoking motives may show little relation to their behaviors and symptoms once they quit smoking.

To achieve the second goal, assessing the discriminant validity of the nicotine dependence questionnaires, we conducted three types of discriminant validity analyses. First, we assessed for scale specificity validity by relating items or scales from nicotine dependence questionnaires to both theoretically related and theoretically remote EMA measures. In general, we found evidence for scale specificity of dependence measures, with dependence scales predicting theoretically related measures but not predicting theoretically unrelated measures, or predicting them less well.

Second, discriminant validity was assessed by using theoretically unrelated dependence scales to predict the target EMA measures that were used in convergent validity analyses. That is, we determined whether a nicotine dependence questionnaire scale or item could predict an EMA outcome (symptom or behavior) that was theoretically relevant to the scale better than could a dependence questionnaire measure that was not conceptually linked with that EMA outcome. Our results showed evidence of discriminant validity in that theoretically relevant scales showed substantially higher levels of predictive accuracy than did the comparison scales (see Table 4). The weight of findings suggests some specificity in terms of the relations between questionnaire dependence measures and behaviors and symptomatic outcomes; that is, dependence measures are not interchangeable in terms of their implications for smokers' behaviors and experiences.

Finally, we assessed method variance contrast discriminant validity by comparing the strength of the relations of our nicotine dependence questionnaire scales with theoretically relevant EMA measures versus the strength of the relations between our nicotine dependence questionnaire scales completed at baseline with other (nondependence) retrospective questionnaire scales completed at baseline. Evidence from these comparisons was mixed. Although all the nicotine dependence measures predicted their theoretically related EMA measures better than they predicted scores on a retrospective self-report measure of alcohol dependence, two of the motives scales (WISDM Negative Reinforcement and WISDM Cognitive Enhancement) showed relatively strong relations with a retrospective self-report measure of depression (see Table 5). However, both of these dependence scales address constructs related to depressive symptoms.

Like all studies, this study has limitations. One limitation is that the nature of the EMA data did not permit more direct tests of some of the predictions that might be deduced from the nicotine dependence measures. For instance, we were not able to determine whether smokers with high-negative reinforcement expectations derived greater distress relief from smoking a given cigarette than did other smokers. In addition, the data were derived from smokers in a cessation study, and such smokers may not be representative of smokers in general.

In conclusion, we found evidence for the convergent validity of many nicotine dependence scales and items. However, the associations between the dependence scales and the EMA measures tended to be modest. This was especially true for the dependence scales assessing dependence motives as opposed to smoking behaviors per se. We also found that the dependence scales and items showed some evidence of discriminant validity. For instance, specific dependence scales and items predicted theoretically related EMA outcomes better than did other dependence measures that had no strong conceptual link to the EMA outcome. There was also evidence that the tested dependence scales predicted theoretically germane EMA responses but not conceptually distinct EMA responses. Given the stringency of the tests (different time frames, different response scales, different response instruments, etc.), these findings provide modest support for the overall construct and discriminant validity of these nicotine dependence scales. However, it is clear the dependence measures do not meet all criteria for discriminant validity. For instance, the intercorrelation between the conceptually distinct dependence measures is much higher than the correlations between the dependence measures and their theoretically linked EMA outcome (e.g., Piper et al., 2004; cf. Campbell & Fiske, 1959). Additional research should be done to assess the construct validity of nicotine dependence measures as they are used to guide important decisions about treatment, genetics, and the nature of psychological disorders. It is not enough for such measures to be significantly related to clinical outcomes. We must also test whether the measures are, in fact, tapping into the theoretical constructs they are intended to measure. In the meantime, we can conclude from this study that smokers sometimes know what we are talking about, but we may often be asking them the wrong questions.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.) Washington, DC: Author.

Baker, T. B., Piper, M. E., McCarthy, D. E., Bolt, D. M., Smith, S. S., Kim, S.-Y., et al. (2007). Time to first cigarette in the morning as an index of ability to quit smoking: Implications for nicotine dependence. *Nicotine and Tobacco Research, 9,* 555–570.

Baker, T. B., Weiss, R. B., Bolt, D. M., von Niederhausern, A., Fiore, M. C., Dunn, D. M., et al. (2009). Human neuronal acetylcholine receptor A5–A3-B4 haplotypes are associated with multiple nicotine dependence phenotypes. *Nicotine and Tobacco Research, 11,* 785–796.

Banthia, R., Malcarne, V. L., Roesch, S. C., Ko, C. M., Greenbergs, H. L., Varni, J. W., & Sadler, G. R. (2006). Correspondence between daily and weekly fatigue reports in breast cancer survivors. *Journal of Behavioral Medicine, 29,* 269–279.

Beirut, L. J., Madden, P. A. F., Breslau, N., Johnson, E. O., Hatsukami, D., Pomerleau, O. F., et al. (2007). Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics, 16,* 24–35.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review, 81,* 506–520.

Bem, D. J., & Funder, D. C. (1978). Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review, 85,* 485–501.

Bolt, D. M., & Johnson, T. R. (in press). Applications of a MIRT model to self-report measures: Addressing score bias and DIF due to individual differences in response style. *Applied Psychological Measurement.*

Boorsboom, D., Mellenbergh, G. J., & Van Heerden, J. V. (2004). The concept of validity. *Psychological Review, 111,* 1061–1071.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Cannon, D. S., Baker, T. B., Piper, M. E., Scholand, M. B., Lawrence, D. L., Drayna, D. T., et al. (2005). Associations between phenylthiocarbamide gene polymorphisms and cigarette smoking. *Nicotine and Tobacco Research, 7,* 853–858.

Chen, X., Zheng, H., Sussman, S., Gong, J., Stacy, A., Xia, J., et al. (2002). Use of the Fagerström Tolerance Questionnaire for measuring nicotine dependence among adolescent smokers in China: A pilot test. *Psychology of Addictive Behaviors, 16,* 260–263.

Cole, D. A., Martin, J. M., Peeke, L., Henderson, A., & Harwell, J. (1998). Validation of depression and anxiety measures in White and Black youths: Multitrait-multimethod analyses. *Psychological Assessment, 10,* 261–276.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous & Mental Disease, 175,* 526–536.

Etter, J. F., Le Houezec, J., & Perneger, T. V. (2003). A self-administered questionnaire to measure dependence on cigarettes: The Cigarette Dependence Scale. *Neuropsychopharmacology, 28,* 359–370.

Fagerström, K. O. (1978). Measuring degree of physical dependence to tobacco smoking with reference to individualization of treatment. *Addictive Behaviors, 3,* 235–241.

Gardner, R. C., & Neufeld, R. W. J. (1987). Use of the change score in correlational analyses. *Educational and Psychological Measurement, 47,* 849–864.

Goedeker, K. C., & Tiffany, S. T. (2008). On the nature of nicotine addiction: A taxometric analysis. *Journal of Abnormal Psychology, 117,* 896–909.

Haberstick, B. C., Timberlake, D., Ehringer, M. A., Lessem, J. M., Hopfer, C. J., Smolen, A., & Hewitt, J. K. (2007). Genes, time to first cigarette and nicotine dependence in a general population sample of young adults. *Addiction, 102,* 655–665.

Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & Fagerström, K.-O. (1991). The Fagerström Test for Nicotine Dependence: A revision of the Fagerström Tolerance Questionnaire. *Addiction, 86,* 1119–1127.

Heiby, E. M. (1995). Assessment of behavioral chaos with a focus on transitions in depression. *Psychological Assessment, 7,* 10–16.

Hudmon, K. S., Gritz, E. R., Clayton, S., & Nisenbaum, R. (1999). Eating orientation, postcessation weight gain, and continued abstinence among female smokers receiving an unsolicited smoking cessation intervention. *Health Psychology, 18,* 29–36.

Hyland, A., Borland, R., Li, Q., Young, H.-H., McNeill, A., Fong, G. T., et al. (2006). Individual-level predictors of cessation behaviours among participants in the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control, 15*(Suppl. 3), iii83–iii94.

Jamison, R. N., Raymond, S. A., Slawsby, E. A., McHugo, G. J., & Baird, J. C. (2006). Pain assessment in patients with low back pain: Comparison of weekly recall and ecological momentary assessment. *Journal of Pain, 7,* 192–199.

Klein, L. C., & Corwin, E. J. (2004). Leptin, hunger, and body weight: Influence of gender, tobacco smoking, and smoking abstinence. *Addictive Behaviors, 29,* 921–927.

Klesges, R. C., Winders, S. E., Meyers, A. W., Eck, L. H., Ward, K. D., Hultquist, C. M., et al. (1997). How much weight gain occurs following smoking cessation? A comparison of weight gain using both continuous and point prevalence abstinence. *Journal of Consulting and Clinical Psychology, 65,* 286–291.

Kozlowski, L. T., Porter, C. Q., Orleans, T., Pope, M. A., & Heatherton, T. (1994). Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND, and HIS. *Drug and Alcohol Dependence, 34,* 211–216.

Lessov, C. N., Martin, N. G., Statham, D. J., Todorov, A. A., Slutske, W. S., Bucholz, K. K., et al. (2004). Defining nicotine dependence for genetic research: Evidence from Australian twins. *Psychological Medicine, 34,* 865–879.

Leventhal, H., Weinman, J., Leventhal, E. A., & Phillips, L. A. (2008). Health psychology: The search for pathways between behavior and health. *Annual Review of Psychology, 59,* 1–29.

McCarthy, D. E., Piasecki, T. M., Fiore, M. C., & Baker, T. B. (2006). Life before and after quitting smoking: An electronic diary study. *Journal of Abnormal Psychology, 115,* 454–466.

Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111,* 172–175.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Mischel, W., & Peake, P. K. (1983). Some facets of consistency: Replies to Epstein, Funder, and Bem. *Psychological Review, 90,* 394–402.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102,* 246–268.

Muthén, B., & Asparouhov, T. (2006). Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors, 31,* 1050–1066.

Payne, T. J., Smith, P. O., McCracken, L. M., McSherry, W. C., & Antony, M. M. (1994). Assessing nicotine dependence: A comparison of the Fagerström Tolerance Questionnaire (FTQ) with the Fagerström Test for Nicotine Dependence in a clinical sample. *Addictive Behaviors, 19,* 307–317.

Pérez-Stable, E. J., Marín, B. V., Marín, G., Brody, D. J., & Benowitz, N. L. (1990). Apparent underreporting of cigarette consumption among Mexican American smokers. *American Journal of Public Health, 80,* 1057–1061.

Piper, M. E., Bolt, D. M., Kim, S.-Y., Japuntich, S. J., Smith, S. S., Niederdeppe, J., et al. (2008). A unique tobacco dependence phenotype suggests primary motives of tobacco dependence. *Journal of Abnormal Psychology, 117,* 747–761.

Piper, M. E., Federman, E. B., McCarthy, D. E., Bolt, D. M., Smith, S. S., Fiore, M. C., et al. (2007). Efficacy of bupropion alone and in combination with nicotine gum. *Nicotine and Tobacco Research, 9,* 947–954.

Piper, M. E., McCarthy, D. E., & Baker, T. B. (2006). Assessing tobacco dependence: A guide to measure evaluation and selection. *Nicotine and Tobacco Research, 8,* 339–351.

Piper, M. E., Piasecki, T. M., Federman, E. B., Bolt, D. M., Smith, S. S., Fiore, M. C., et al. (2004). A multiple motives approach to tobacco dependence: The Wisconsin Inventory of Smoking Dependence Motives (WISDM-68). *Journal of Consulting and Clinical Psychology, 72,* 139–154.

Prokhorov, A. V., De Moor, C., Pallonen, U. E., Hudmon, K. S., Koehly, L., & Hu, S. (2000). Validation of the modified Fagerström Tolerance Questionnaire with salivary cotinine among adolescents. *Addictive Behaviors, 25,* 429–433.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1,* 385–401.

Saccone, S. F., Hinrichs, A. L., Saccone, N. L., Chase, G. A., Konvicka, K., Madden, P., et al. (2007). Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3,713 SNPs. *Human Molecular Genetics, 16,* 36–49.

Selzer, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic intrument. *American Journal of Psychiatry, 127,* 1653–1658.

Shiffman, S., Hickcox, M., Paty, J., Gnys, M., Kassel, J., & Richards, T. J. (1997). The abstinence-violation effect following smoking lapses and temptations. *Cognitive Research and Therapy, 21,* 497–523.

Shiffman, S., Patten, C., Gwaltney, C., Paty, J., Gnys, M., Kassel, J., et al. (2006). Natural history of nicotine withdrawal. *Addiction, 101,* 1822–1832.

Shiffman, S., & Sayette, M. A. (2005). Validation of the nicotine dependence syndrome scale (NDSS): A criterion-group design contrasting chippers and regular smokers. *Drug and Alcohol Dependence, 79,* 45–52.

Shiffman, S., Waters, A. J., & Hickcox, M. (2004). The Nicotine Dependence Syndrome Scale: A multidimensional measure of nicotine dependence. *Nicotine and Tobacco Research, 6,* 327–348.

Skinner, H. A. (1979). A multivariate evaluation of the MAST. *Journal of Studies on Alcohol, 40,* 831–844.

Smith, R. E., Leffingwell, T. R., & Ptacek, J. T. (1999). Can people remember how they coped? Factors associated with discordance between same-day and retrospective reports. *Journal of Personality and Social Psychology, 76,* 1050–1061.

Stone, A. A., Schwartz, J. E., Neale, J. M., Shiffman, S., Marco, C. A., Hickcox, M., et al. (1998). A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology, 74,* 1670–1680.

Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 26–39). New York: Russell Sage Foundation.

Todd, M., Armeli, S., Tennen, H., Carney, M. A., Ball, S. A., Kranzler, H. R., & Affleck, G. (2005). Drinking to cope: A comparison of questionnaire and electronic diary reports. *Journal of Studies on Alcohol, 66,* 121–129.

Todd, M., Tennen, H., Carney, M. A., Armeli, S., & Affleck, G. (2004). Do we know how we cope? Relating daily coping reports to global and time-limited retrospective assessments. *Journal of Personality and Social Psychology, 86,* 310–319.

Welsch, S. K., Smith, S. S., Wetter, D. W., Jorenby, D. E., Fiore, M. C., & Baker, T. B. (1999). Development and validation of the Wisconsin Smoking Withdrawal Scale. *Experimental and Clinical Psychopharmacology, 7,* 354–361.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment.* Malabar, FL: Krieger Publishing Company.

Received December 23, 2008
Revision received June 19, 2009
Accepted June 22, 2009 ■