Published in final edited form as:

Psychol Assess. 2009 December ; 21(4): 595–607. doi:10.1037/a0017312.

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

S. J. Japuntich

Center for Tobacco Research and Intervention, University of Wisconsin School of Medicine and Public Health

M. E. Piper

Center for Tobacco Research and Intervention, University of Wisconsin School of Medicine and Public Health

T. R. Schlam

Center for Tobacco Research and Intervention, University of Wisconsin School of Medicine and Public Health

D. M. Bolt

Department of Educational Psychology, University of Wisconsin-Madison

T. B. Baker

Center for Tobacco Research and Intervention, University of Wisconsin School of Medicine and Public Health

Abstract

Few studies have examined whether nicotine dependence self-report questionnaires can predict specific behaviors and symptoms at specific points in time. The current study used data from a randomized clinical trial ($n = 608$; Piper et al., 2007) to assess the construct validity of scales and items from three nicotine dependence measures: the Fagerström Test for Nicotine Dependence (Heatherton, Kozlowski, Frecker & Fagerström, 1991), the Nicotine Dependence Syndrome Scale (Shiffman, Waters & Hickcox, 2004), and the Wisconsin Index of Smoking Dependence Motives (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 et al., 1991). Researchers administer self-report items 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; Beirut et al., 2007; Cannon et al., 2005; Etter, Le Houezec, & Perneger, 2003; Haberstick, Timberlake, Ehringer, Lessem, Hopfer, Smolen & Hewitt, 2007; Hyland et al., 2006; Kozlowski, Porter, Orleans, Pope, & Heatherton, 1994; Saccone et al., 2007). In addition, researchers use self-report items 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, & Nisenbaum, 1999; Muthen & Asparouhov, 2006; Piper et al., 2008; Shiffman et al., 2004). Researchers also use self-report items 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 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 current study, we assess 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 amongst variables. According to Campbell and Fiske (1959), the construct validation of a measure can be inferred based on 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

¹Recently there has been a debate in the measurement literature about whether constructs should be conceptualized via nomological networks. Boorsboom, Mellenbergh, and Van Heerden (2004) argue that measures gain validity to the extent that one can prove that (a) the construct exists and is worth measuring and (b) that changes in the construct cause changes in the measure of that construct. Thus, Boorsboom et al., view 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.

²Convergent 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).

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 the different experiences use a similar measurement/assessment strategy (e.g., use of a questionnaire as a data collection strategy; Campbell & Fiske, 1959). Campbell and Fiske (1959) posit multiple types of comparisons that provide evidence for the discriminant validity of measures, two of which will be 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 will 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; Heatherington et al., 1991) or with DSM criteria (DSM-IV-TR; American Psychiatric Association, 2000) or by using the measure to predict relapse likelihood (predictive validity; e.g., Piper et al., 2004). Discriminant validity is rarely assessed.

Correlating one dependence questionnaire with another provides information on the clinical utility of the questionnaire (i.e., they reveal agreement with other measures that are used for clinical diagnosis and permit the user to gauge relapse likelihood). However, these relations do not provide much theoretical insight. These relations 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., 2005). If this measure is then found to predict the tendency to relapse back to smoking, it would be tempting to speculate that the individual relapses quickly because s/he experiences especially strong withdrawal symptoms if s/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, are high in neuroticism, or who believe that they are highly dependent, might rate all sorts of dependence-related symptoms quite highly. Such individuals might also be more likely to relapse.

The proposed research tests the ability of different questionnaire measures of nicotine dependence to assess relatively specific behavioral outcomes collected via real-time data acquisition. This strategy was selected for two reasons. First, evidence attests to the validity of real-time data acquisition methods; e.g., 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 amongst measures that are relatively uncontaminated by method variance (Campbell & Fiske, 1959; Cole, Martin, Peeke, Henderson, & Harwell, 1998). As compared to 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.

As we have argued earlier, clinicians and theorists make decisions and inferences based on 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, Tennen, Carney, Armili, & Affleck, 2004; Todd, Armeli, Tennen, Carney, Ball, Kranzler, & Affleck, 2005), and sometimes well (Pain: Jamison, Raymond, Slawsby, McHugo, & Baird, 2006; Fatigue: Banthia, Malcarne, Roesch, Ko, Greenbergs, Varni, & Sadler, 2006) with real-time behavioral ratings. This variability may be due to the fact that 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 by 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, Vanoss Marín, Marín, Brody, & Benowitz, 1990; Prokhorov, De Moor, Pallonen, Hudmon, Koehly, & Hu, 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.

The current study evaluates the construct validity of three nicotine dependence measures 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): (1) the convergent validity of a dependence scale or item (i.e., whether a dependence measure predicts a theoretically-related EMA symptom or behavior); (2) 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 (3) the method variance contrasts of a dependence item or scale (i.e., the extent to which a self-report dependence measure predicts a theoretically-related EMA outcome better than it predicts an unrelated self-report measure).

Method

Participants—Participants were 608 smokers (57.9% female, 76% Caucasian; see Table 1 for demographics), recruited as part of a clinical trial (Piper et al., 2007) conducted in

Milwaukee, WI. 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 mental health issues 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 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 minute), 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 pre-quit, on the quit day, and 1 week post-quit. From 1 week pre-quit to 1 week post-quit, 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 post-quit.

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 minutes (Csikszentmihalyi & Larson, 1987; Shiffman, Hickcoz, Paty, Gnys, Kassel, & Richards, 1997). In the current study, participants carried cell phones from one week pre-quit to one week post-quit. The data were collected via an interactive voice recording (IVR) system. When we examined smoking at the first call of the day, we excluded participants who reported smoking greater than 10 cigarettes between the evening call and the morning call (thus the top 5% of respondents were excluded from only these analyses) because we felt this degree of smoking during the night was unlikely and therefore the participants had likely misunderstood the question.

Measures

Center for Epidemiologic Studies Depression Scale—The Center for Epidemiologic Studies Depression Scale (CES-D, 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.

Fagerström Test of Nicotine Dependence—The Fagerström Test of Nicotine Dependence (FTND, Heatherton et al., 1991) comprises 6 items. Scores on the scale range from 0 to 10, with higher numbers indicating greater dependence.

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”).

Nicotine Dependence Syndrome Scale—The Nicotine Dependence Syndrome Scale (NDSS, Shiffman et al., 2004) is a 23-item self-report measure. It has 5 scales: tolerance, drive, stereotypy, continuity and priority.

Wisconsin Inventory of Smoking Dependence Motives—The Wisconsin Inventory of Smoking Dependence Motives (WISDM, Piper et al., 2004) is a 68-item measure which 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.

Wisconsin Smoking Withdrawal Scale—The Wisconsin Smoking Withdrawal Scale (WSWS, Welsch, Smith, Wetter, Jorenby, Fiore, & Baker, 1999) is a 28-item measure of nicotine withdrawal. It comprises seven scales: Anger, Anxiety, Concentration, Craving, Hunger, Sadness, and Sleep. This study used a condensed version of the WSWS 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

Our analyses do not constitute a full multitrait-multimethod 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, we will not complete the evaluations for that side of the matrix. We also do not have two identical sets of trait by method comparisons, thus we use different questionnaires for the method variance contrasts and scale specificity analyses and for the convergent validity comparisons. Another caveat is that the different questionnaire-EMA facets reflect not only 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 we focused on analytic comparisons that address the validity of the questionnaire nicotine dependence measures, and used the MTMM model (Campbell and 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. We were able to test the construct validity of only those global self-report scales 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; i.e., 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, 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 pre- to post-quit increases on the relevant EMA measures (i.e., difference scores between pre-quit and post-quit levels on the EMA measure).³ It is important to note that this strategy is conservative since 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: e.g., 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 pre-quit, 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 pre-cessation smoking heaviness in the week after participants completed the FTND questionnaire).

Control Variables—Linear regression was used for all analyses, and gender and treatment condition were entered as covariates. Analyses of dependence motives also controlled for level of dependence (measured via cigarettes per day)⁴; 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 timepoint 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; Klesges et al., 1997; Klein & Corwin, 2004). For other tests of specific behavioral manifestations of dependence, cigarettes per day 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 Heatherton et al., 1991; Baker et al., 2007; Lessov et al., 2004).

³An 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 pre-cigarette to post-cigarette). However, the temporal resolution of EMA reports was not ideal for this sort of analysis. This would have required pre/post ratings of individual cigarettes and because EMA data were collected only four times/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 or not a person had just smoked) because such simultaneous ratings raise concerns about reciprocal causality.

⁴When analyses of dependence motives were conducted without controlling for cigarettes per day, the pattern of significant effects remained the same.

Time Coding—The pre-quit period was defined a priori as 7 to 4 days pre-quit, 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 due to changes in smoking behavior and affect that might have occurred immediately prior to a quit attempt. The post-quit period was defined a priori as 1 to 3 days post-quit, 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 post-quit were not included because withdrawal and symptomatic rebound were expected to be most intense during the first few days of the quit attempt (Shiffman & Sayette, 2005; Shiffman, et al., 2006) and later time points were likely to be more affected by post-quit smoking as participants relapsed.⁵ The pre-quit assessments resulted in 5536 EMA reports; the post-quit assessments resulted in 5077 reports.

Procedure—Control variables were added to the model in the first step. In the next step, the dependence scale/item of interest was entered as the predictor variable. Convergent validity was conceptualized as the increment in variance accounted for in the dependent variable by the independent 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

We examined two types of discriminant validity 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, we selected EMA measures such that one measure was highly relevant to the dependence questionnaire measure while the *contrast* EMA measure was highly dissimilar. For example, for the dependence motives scales, we picked the WSWs scale (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. The smoking behavior EMA measures tended to be count measures (e.g., cigarettes per day). Thus, we chose the only non-smoking-related categorical measure in our 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).

In addition to evaluating whether the nicotine dependence measure was related to a theoretically unrelated EMA measure, we also assessed 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, we 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

⁵Because 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, 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.

questionnaire measure. We chose four comparison nicotine dependence questionnaire scales that were theoretically distinct from the dependence questionnaire scales used in the convergent validity analyses. This decision was made because 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, we chose to use nicotine dependence measures because their response scales were identical to the target nicotine dependence scales. The following dependence scales were judged 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 unrelated comparison dependence questionnaire scale/item. We assessed EMA scale specificity 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 if 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, we used two theoretically unrelated comparison questionnaire measures: 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 is a yes-no checklist, 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 to 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—In general, 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)⁶. 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 post-quit. Similarly, the WISDM Negative Reinforcement scale (e.g., “Smoking a cigarette improves my mood”) predicted emergent negative affect post-quit. The craving scales, however, did not perform as well. The WISDM Craving scale did not predict emergent craving post-quit⁷. The NDSS Drive scale preformed better than the WISDM Craving scale, but still predicted pre-post increases in craving only marginally. The WISDM Weight Control scale predicted neither post-quit changes in hunger in the first several days post-quit nor weight gain at 4 or 8 weeks post-quit. Moreover, even when relations between predictors and theoretically-related EMA

⁶We also examined whether the dependence motives measures would predict theoretically-related symptoms pre-quit and post-quit, in addition to predicting the increase from pre- to post-quit. All of the motives measures predicted theoretically-related pre-quit and post-quit symptoms (even the cravings scales). However, since the dependence motives measures assess whether smoking helps to control certain symptoms, we felt that the post-quit increase in such symptoms was a more relevant and demanding test.

⁷This lack of prediction was not due to a lack of variance in the craving difference scores caused by high relations between pre- and post-quit craving. Pre- and post-quit craving ratings were correlated, but not as strongly as were pre-post rating of the other EMA variables.

outcomes were significant, the amount of variance accounted for by the scales was small (1–2%; see Table 2).⁸

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 (2–37%, see Table 2). FTND item 1 (“How soon after waking do you smoke your first cigarette?”) predicted the EMA pre-quit 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 pre-quit 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 pre-quit 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 current 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 pre- to post-quit increase on a theoretically-related EMA measure. First, we computed this analysis for the variable pair with the smallest, significant relation with one another (see Table 2): global NDSS Drive and EMA-assessed WSWs Craving. Of those who scored in the top 25% on the NDSS Drive scale, only 31.8% scored in the top 25% on the WSWs Craving scale. Next we computed this analysis for a variable pair with a moderate level of association: the WISDM Cognitive Enhancement scale and the WSWs “trouble concentrating” item. Of those who scored in the top 25% on the WISDM scale, 37.6% scored in the top 25 percentile on pre- to post-quit difference scores on the EMA item “had trouble concentrating.” 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 pre-quit (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.

Scale Specificity Discriminant Validity

Dependence-measures—Our next goal was to assess the scale specificity of the nicotine dependence questionnaire measures. As expected, none of the motives scales (WISDM Cognitive Enhancement, WISDM Negative Reinforcement and NDSS Drive) predicted the EMA measure of post-quit increases in hunger (see Table 3). Similarly, 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, while these dependence measures predicted EMA items from theoretically relevant domains, they did not predict EMA items from conceptually distinct domains.

⁸To examine the upper limit of these types of comparisons, we compared WSWs ratings during the first 3 days post-quit with retrospective WSWs ratings at the week 1 visit 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 post-quit; 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. This suggests that a single global measure could predict meaningful amounts of variance in the phasic EMA measures; i.e., variation in phasic EMA measures over time did not preclude meaningful associations.

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, and NDSS Drive scales had stronger relations with their relevant EMA measures than did theoretically-unrelated comparison questionnaire measures. There was also substantial evidence of discriminant validity with regards to the smoking-targeted measures. For instance, the EMA measure of cigarettes smoked per day was highly related to the FTND cigarettes per day item and to the two tolerance scales (NDSS; WISDM). This EMA item (cigarettes per day) was also significantly related to the WISDM Taste/Sensory Properties scale, WISDM Social/Environmental Goads scale, and NDSS Continuity scale, but the relations were much weaker. Morning smoking was only predicted by FTND item 1, consistent with the behavior targeted by that item.

Method Contrasts Discriminant Validity

Motives Scales—The motives scales showed mixed support for the method contrasts discriminant validity. WISDM Cognitive Enhancement was more strongly related to EMA reports of increased difficulty concentrating post-quit than it was to a questionnaire measure of alcohol dependence. However, it was about equally as strongly related to retrospective self-reports of depression symptoms, a questionnaire with a more similar response scale than the alcohol dependence measure. The NDSS Drive scale was more strongly related to EMA reports of post-quit increases in craving than it was to the questionnaire measures of alcohol dependence or depression. The WISDM Negative Reinforcement scale was more strongly related to post-quit 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 may be a non-optimal comparison scale. The cigarette use scales fared better 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.

Discussion

The current study had two main goals: 1) 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 2) 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 relevant behaviors and symptoms was quite modest: e.g., the percentage of variance in EMA-measured emergent withdrawal symptoms accounted for by the relevant questionnaire dependence measures was quite low (1–2% of the variance). The dependence measures targeting smoking behavior performed somewhat better (accounting for some 2–37% 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 that the scales do not permit accurate prediction of which individuals will be extreme on relevant behaviors or symptoms during daily smoking or during quit attempts.

There are many plausible explanations for the modest predictive relations between the dependence scales/items and the EMA reports. First, the two types of measures used different response scales and different timeframes for responding. 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. Participants' nicotine dependence questionnaire responses may reflect trait-like characteristics that are independent of nicotine dependence (e.g., neuroticism) and that do not affect EMA ratings similarly. 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/post-cessation effects. Fourth, the EMA items elicited responses about particular episodes while the dependence scales/items elicited global, trait-like 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, trait-like ratings because individuals may behave differently in different situations (Mischel & Shoda, 1995). It is also possible that individuals differ in their ability to predict their behaviors or symptoms, such that more impressive associations would be found in a subgroup of individuals (Bem & Allen, 1974; Bem & Funder, 1978; Mischel & Shoda, 1995). Finally, it may be that some dependence measures are intrinsically poorly related to abstinence induced change. This may be because smokers may have little idea of the extent to which smoking actually controls such symptoms as negative affect or craving. Therefore, their ratings of the extent to which they depend on tobacco for these purposes are poorly related to abstinence induced change. It is also possible that some of the constructs targeted by dependence measures are intrinsically poorly related to abstinence induced change.

Researchers have reported modest relations between global self-report measures and specific real-time measures in diverse research areas (e.g., coping: Todd et al., 2004; Stone et al., 1998; Smith, et al., 1999; 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 to daily reports during the same period. The current study also showed impressive relations between global, retrospective self-report and an amalgam of the phasic EMA reports (see footnote 8 where retrospective recall at 1-week post-quit was meaningfully related to the phasic EMA symptoms gathered over 1–3 days postquit). This suggests that there is enough reliable variance in the phasic EMA measures to permit their accurate prediction. However, the subjects in this research either did not know what their withdrawal would be like, or some dependence questions elicited responses that were 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) construct. 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 were assessing specific symptoms (Fatigue: Banthia et al., 2006; Pain: Jamison et al., 2006), whereas the studies that found less clear relationships were assessing coping (e.g., coping: Todd, Tennen, Carney, Armili, & Affleck, 2004; Stone et al., 1998; Smith, Leffingwell, & Ptacek, 1999; and drinking to cope: Todd, Armeli, Tennen, Carney, Ball, Kranzler, & Affleck, 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; i.e., they did not entail extrapolation from prequit measures to postquit events.

While the associations between some dependence questionnaire scales and EMA symptom ratings were modest, it is true that some of the dependence scales and items are significantly related to theoretically relevant symptoms and behaviors recorded in real time. This finding provides some support for theoretical inferences made on the basis of some of the scales (e.g., negative reinforcement). However, the level of associations suggest 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 raises questions about the usefulness of some of the information that counselors extract in counseling sessions with smokers because smokers' characterization of their smoking motives may show little relation to behaviors and symptoms that occur 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.

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. While some of the comparison scales were related to the EMA measure of cigarettes smoked per day (Table 4), the associations with the theoretically relevant scales were much higher. The weight of findings suggests some specificity in terms of the relations between questionnaire dependence measures and behaviors and symptomatic outcomes; i.e., dependence measures are not interchangeable in terms of their implications for smokers' behaviors and experiences.

Finally, we assessed for method variance contrast discriminant validity by comparing the strength of the relations of our nicotine dependence questionnaire scales with EMA measures versus the strength of the relations between our nicotine dependence questionnaire scales with other retrospective questionnaire scales. Evidence for these comparisons was mixed. While 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) were more strongly related to a retrospective self-report measure of depression than to a theoretically-related EMA measure. However, both of these dependence scales address constructs related to depressive symptoms.

Limitations and Conclusions

Like all studies, this study has certain 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 reinforcement expectations derived greater distress relief from smoking a given cigarette than did other smokers.

In conclusion, we found evidence for the convergent validity of many nicotine dependence scales and items. However, the associations between the dependence scales and 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, they predicted theoretically related EMA outcomes better than did 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.), this provides modest support for the overall construct and discriminant validity of these nicotine dependence scales. However, it is clear that 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're talking about, but we may, in fact, be asking them the wrong questions.

Acknowledgments

This research was supported by National Institutes of Health grants CA84724-05 and DA0197-06.

References

- American Psychiatric Association. (DSM-IV-TR) Diagnostic and Statistical Manual of Mental Disorders. 4th edition, text revision. American Psychiatric Press, Inc.; Washington, DC: 2000.
- Baker TB, Piper ME, McCarthy DE, Bolt DM, Smith SS, Kim S-Y, et al. Time to first cigarette in the morning as an index of ability to quit smoking: Implications for nicotine dependence. *Nicotine and Tobacco Research* 2007;9S4:555–570.
- Banthia R, Malcarne VL, Roesch SC, Ko CM, Greenbergs HL, Varni JW, Sadler GR. Correspondence between daily and weekly fatigue reports in breast cancer survivors. *Journal of Behavioral Medicine* 2006;29:269–279. [PubMed: 16724282]
- Beirut LJ, Madden PAF, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF, Swan GE, et al. Novel genes identified in a high-density genome wide association study for nicotine dependence. *Human Molecular Genetics* 2007;2007:16, 24–35.
- Bem DJ, Allen A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review* 1974;81:506–520.
- Bem DJ, Funder DC. Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review* 1978;85:485–501.
- Boorsboom D, Mellenbergh GJ, Heerden JV. The concept of validity. *Psychological Review* 2004;111:1061–1071. [PubMed: 15482073]
- Bolt, DM.; Johnson, TR. Applications of a MIRT model to self-report measures: Addressing score bias and DIF due to individual differences in response style. in press *Applied Psychological Measurement*
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait—multimethod matrix. *Psychological Bulletin* 1959;56:81–105. [PubMed: 13634291]

- Cannon DS, Baker TB, Piper ME, Scholand MB, Lawrence DL, Drayna DT, et al. Associations between phenylthiocarbamide gene polymorphisms and cigarette smoking. *Nicotine and Tobacco Research* 2005;7:853–858. [PubMed: 16298720]
- Chen X, Zheng H, Sussman S, Gong J, Stacy A, Xia J, et al. Use of the Fagerström tolerance questionnaire for measuring nicotine dependence among adolescent smokers in China: A pilot test. *Psychology of Addictive Behaviors* 2002;16:260–263. [PubMed: 12236462]
- Cole DA, Martin JM, Peeke L, Henderson A, Harwell J. Validation of depression and anxiety measures in white and black youths: Multitrait-multimethod analyses. *Psychological Assessment* 1998;10:261–276.
- Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychological Bulletin* 1955;52:281–302. [PubMed: 13245896]
- Csikszentimihalyi M, Larson R. Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease* 1987;175:526–536. [PubMed: 3655778]
- Etter JF, Le Houezec J, Perneger TV. A self-administered questionnaire to measure dependence on cigarettes: The Cigarette Dependence Scale. *Neuropsychopharmacology* 2003;28:359–370. [PubMed: 12589389]
- Goedeker KC, Tiffany ST. On the nature of nicotine addiction: A taxometric analysis. *Journal of Abnormal Psychology* 2008;117:896–909. [PubMed: 19025235]
- Haberstick BC, Timberlake D, Ehringer MA, Lessem JM, Hopfer CJ, Smolen A, Hewitt JK. Genes, time to first cigarette and nicotine dependence in a general population sample of young adults. *Addiction* 2007;102:655–665. [PubMed: 17309537]
- Heatherton TF, Kozlowski LT, Frecker, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *British Journal of Addiction* 1991;86:1119–1127. [PubMed: 1932883]
- Hudmon KS, Gritz ER, Clayton S, Nisenbaum R. Eating orientation, postcessation weight gain, and continued abstinence among female smokers receiving an unsolicited smoking cessation intervention. *Health Psychology* 1999;18:29–36. [PubMed: 9925043]
- Hyland A, Borland R, Li Q, Young H-H, McNeill A, Fong GT, et al. Individual-level predictors of cessation behaviours among participants in the International Tobacco Control (ITC) Four Country Survey. *Tobacco Control* 2006;15(Suppl III):iii83–iii94. [PubMed: 16754952]
- Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and ecological momentary assessment. *Journal of Pain* 2006;7:192–199. [PubMed: 16516825]
- Klein LC, Corwin EJ. Leptin, hunger, and body weight: Influence of gender, tobacco smoking, and smoking abstinence. *Addictive Behaviors* 2004;29:921–927. [PubMed: 15219336]
- Klesges RC, Winders SE, Meyers AW, Eck LH, Ward KD, Hultquist CM, et al. 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* 1997;65:286–291. [PubMed: 9086692]
- Kozlowski LT, Porter CQ, Orleans T, Pope MA, Heatherton T. Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND, and HIS. *Drug and Alcohol Dependence* 1994;34:211–216. [PubMed: 8033758]
- Lesso CN, Martin NG, Statham DJ, Todorov AA, Slutske WS, Bucholz KK, Heath AC, Madden PA. Defining nicotine dependence for genetic research: Evidence from Australian twins. *Psychological Medicine* 2004;34:865–79. [PubMed: 15500307]
- Messick S. Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist* 1995;50:741–749.
- Mischel W, Peake PK. Some Facets of Consistency: Replies to Epstein, Funder, and Bem. *Psychological Review* 1983;90(4):394–402.
- Mischel W, Shoda Y. A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review* 1995;102:246–268. [PubMed: 7740090]
- Muthen B, Asparouhov T. Item response mixture modeling: application to tobacco dependence criteria. *Addictive Behaviors* 2006;31:1050–66. [PubMed: 16675147]

- Pérez-Stable EJ, Vanoss Marín B, Marín G, Brody DJ, Benowitz NL. Apparent underreporting of cigarette consumption among Mexican American smokers. *American Journal of Public Health* 1990;80:1057–1061. [PubMed: 2382741]
- Piper ME, Bolt DM, Kim S-Y, Japuntich SJ, Smith SS, Niederdeppe J, et al. A unique tobacco dependence phenotype suggests primary motives of tobacco dependence. *Journal of Abnormal Psychology* 2008;117:747–761. [PubMed: 19025223]
- Piper ME, Federman EB, McCarthy DE, Bolt DM, Smith SS, Fiore MC, et al. Efficacy of bupropion alone and in combination with nicotine gum. *Nicotine and Tobacco Research* 2007;9:947–954. [PubMed: 17763111]
- Piper ME, McCarthy DE, Baker TB. Assessing tobacco dependence: A guide to measure evaluation and selection. *Nicotine and Tobacco Research* 2006;8:339–351. [PubMed: 16801292]
- Piper ME, Piasecki TM, Federman EB, Bolt DM, Smith SS, Fiore MC, et al. A Multiple Motives Approach to Tobacco Dependence: The Wisconsin Inventory of Smoking Dependence Motives (WISDM-68). *Journal of Consulting and Clinical Psychology* 2004;72:139–154. [PubMed: 15065950]
- Prokhorov AV, De Moor C, Pallonen UE, Hudmon KS, Koehly L, Hu S. Validation of the modified Fagerström Tolerance Questionnaire with salivary cotinine among adolescents. *Addictive Behaviors* 2000;25:429–33. [PubMed: 10890296]
- Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement* 1977;1:385–401.
- Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden P, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3,713 SNPs. *Human Molecular Genetics* 2007;16:36–49. [PubMed: 17135278]
- Selzer ML. The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry* 1971;127:1653–1658. [PubMed: 5565851]
- Shiffman S. Assessing smoking patterns and motives. *Journal of Consulting & Psychology* 1993;61:732–742.
- Shiffman S, Hickcox M, Paty J, Gnys M, Kassel J, Richards TJ. The Abstinence-violation effect following smoking lapses and temptations. *Cognitive Research and Therapy* 1997;21:497–523.
- Shiffman S, Patten C, Gwaltney C, Paty J, Gnys M, Kassel J, et al. Natural history of nicotine withdrawal. *Addiction* 2006;101:1822–1832. [PubMed: 17156182]
- Shiffman S, Sayette MA. Validation of the nicotine dependence syndrome scale (NDSS): a criterion-group design contrasting chippers and regular smokers. *Drug and Alcohol Dependence* 2005;79:45–52. [PubMed: 15943943]
- Shiffman S, Waters AJ, Hickcox M. The Nicotine Dependence Syndrome Scale: A multidimensional measure of nicotine dependence. *Nicotine and Tobacco Research* 2004;6:327–348. [PubMed: 15203807]
- Skinner HA. A multivariate evaluation of the MAST. *Journal of Studies on Alcohol* 1979;40:831–844. [PubMed: 513775]
- Smith RE, Leffingwell TR, Pracek JT. Can people remember how they coped? Factors associated with discordance between same-day and retrospective reports. *Journal of Personality and Social Psychology* 1999;76:1050–1061.
- Stone AA, Schwartz JE, Neale JM, Shiffman S, Marco CA, Hickcox M, et al. A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology* 1998;74:1670–1680. [PubMed: 9654765]
- Stone, AA.; Shiffman, SS.; DeVries, MW. Ecological Momentary Assessment. In: Kahneman, D.; Diener, E.; Schwarz, N., editors. *Well-being: The foundations of hedonic psychology*. Russell Sage Foundation; New York: 1999. p. 26-39.
- Todd M, Armeli S, Tennen H, Carney MA, Ball SA, Kranzler HR, Affleck G. Drinking to cope: A comparison of questionnaire and electronic diary reports. *Journal of Studies on Alcohol* 2005;66:121–129. [PubMed: 15830912]

- Todd M, Tennen H, Carney MA, Armeli S, Affleck G. Do we know how we cope? Relating daily coping reports to global and time-limited retrospective assessments. *Journal of Personality and Social Psychology* 2004;86:310–319. [PubMed: 14769086]
- Welsch SK, Smith SS, Wetter DW, Jorenby DE, Fiore MC, Baker TB. Development and validation of the Wisconsin Smoking Withdrawal Scale. *Experimental and Clinical Psychopharmacology* 1999;7:354–361. [PubMed: 10609970]
- Wiggins, JS. *Personality and Prediction: Principles of personality assessment*. Krieger Pub Co; Malabar, FL: 1973.

Dependence Questionnaire Measure**EMA Measure****Convergent Validity**

Dependence measure

Theoretically Relevant Target Measure

Scale→EMA Specificity

Dependence measure

Target (relevant) EMA measure

Contrast EMA measure

EMA→Scale Specificity

Convergent Dependence measure

Convergent EMA measure

Contrast Dependence measure

Method Variance Contrast

Convergent Dependence measure

Convergent EMA measure

Contrast Nondependence measure

Figure 1.

Overall analytical approach for construct validation of nicotine dependence scales. (1) Convergent validity was tested by assessing the strength of the relation between a target nicotine dependence scale and a theoretically related EMA measure (2) Discriminant validity was measured 3 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 to 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.

Table 1

Demographics

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)

Table 2

Convergent Validity of Nicotine Dependence Measures with Theoretically-Related Outcomes (Using Linear Regression Controlling for Gender and Treatment)

Dependence Measure	EMA Measure	Timepoint	Δr^2 *	β	t	p
WISDM Cog. Enhance.	WSWS Conc.	Post-pre**	.02	.14	3.25	.001
WISDM Crav.	WSWS Crav.	Post-pre	.001	.03	.72	.47
NDSS Drive	WSWS Crav.	Post-pre	.01	.09	1.97	.05
WISDM Neg. Rein.	WSWS Neg. Affect	Post-pre	.01	.12	2.79	.01
WISDM Wt. Cont.	WSWS hung.	Post-pre	.003	.05	1.14	.26
	Δ in weight since bsn.	Post-quit 4 weeks	.001	.03	.65	.51
	Δ in weight since bsn.	Post-quit 8 weeks	.002	-.05	-.92	.36
FTND Item 1	# of cigs since last call	Pre-quit morning	.03	.17	3.76	<.001
	# of cigs since last call	Post-quit morning	.02	.15	3.43	.001
FTND Item 4	Cigs. smoked since yesterday	Pre-quit	.37	.62	18.17	<.001
NDSS Tol.	Cigs. smoked since yesterday	Pre-quit	.09	.31	7.71	<.001
WISDM Tol.	Cigs smoked since yesterday	Pre-quit	.10	.32	7.96	<.001

* Δr^2 denotes the change in r^2 between the model with only control variables to the model with the nicotine dependence measure added.

** Post-pre indicates the difference between the post-quit scores and the pre-quit scores.

Table 3

Scale Specificity Discriminative Validity for the Dependence Measures

Dependence Measure	EMA	Timepoint	Δr^2 *	B	t	p
WISDM Cog. Enhanc.	WSWS Conc.	Post-pre**	.02	.14	3.25	.001
	WSWS hung.	Post-pre	.003	.06	1.27	.21
WISDM Neg. Rein.	WSWS Neg. Affect	Post-pre	.01	.12	2.79	.01
	WSWS hung.	Post-pre	.003	.05	1.20	.23
NDSS Drive	WSWS Crav.	Post-pre	.008	.09	1.97	.05
	WSWS hung.	Post-pre	.004	.06	1.48	.14
WISDM Tol.	Cigs. smoked since yesterday	Pre-quit	.10	.32	7.96	<.001
	Stressful sit. since last call	Pre-quit	.003	-.06	-1.23	.22
NDSS Tol.	Cigs. smoked since yesterday	Pre-quit	.09	.31	7.71	<.001
	Stressful sit. since last call	Pre-quit	.001	-.02	-0.43	.67
FTND Item 1	# of cigs since last call	Pre-quit morning	.03	.17	3.76	<.001
	Stressful event since last call	Pre-quit morning	.00	-.01	-.19	.85
FTND item 4	Cigs. smoked since yesterday	Pre-quit	.37	.62	18.17	<.001
	Stressful sit. since last call	Pre-quit	.003	-.08	-1.22	.22

* Δr^2 denotes the change in r^2 between the model with only control variables to the model with the nicotine dependence measure added.

** Post-pre indicates the difference between the post-quit scores and the pre-quit scores.

Table 4

Scale Specificity Discriminant Validity for the EMA Measures

EMA Dependent Variable	Dependence Measure (Independent Variable)	Δr^2 *	B	t	p
WSWS Concentration (post - pre-quit)	WISDM Cog. Enhanc.	.02	.14	3.25	.001
	WISDM Soc./envr. Goads	.00	-.02	-.53	.60
	WISDM Taste	.00	.02	.55	.58
	NDSS Cont.	.00	.003	.06	.95
WSWS Crav. (post- pre quit)	NDSS Drive	.01	.09	1.97	.05
	WISDM Soc./envr. Goads	.01	-.07	-1.59	.11
	WISDM Taste	.001	.01	.30	.77
	NDSS Cont.	.001	-.02	-.36	.72
WSWS neg. affect	WISDM Neg. Rein.	.01	.12	2.79	.006
	WISDM Soc./envr. Goads	.00	-.06	-1.40	.16
	WISDM Taste	.01	.07	1.57	.12
	NDSS Cont.	.002	.04	.89	.37
Cigs per day	WISDM Tol.	.10	.32	7.96	<.001
	NDSS Tol.	.09	.31	7.71	<.001
	FTND item 4	.37	.62	18.17	<.001
	WISDM Soc./envr. Goads	.01	.09	2.21	.03
Cigs smoked since last call (morning call pre-quit)	WISDM Taste	.02	.13	3.01	.003
	NDSS Cont.	.02	.13	3.08	.002
	FTND Item 1	.03	.17	3.76	<.001
	WISDM Soc./envr. Goads	.00	.01	.10	.92
	WISDM Taste	.003	.06	1.23	.22
	NDSS Cont.	.002	.04	.99	.32

* Δr^2 denotes the change in r^2 between the model with only control variables to the model with the nicotine dependence measure added.

Table 5

Method Variance Contrast Discriminant Validity

Dependence Measure	Comparison Measure	Δr^2 *	β	t	p
WISDM Cog. Enhanc.	EMA WSWWS Conc. post-pre	.02	.14	3.25	.001
	MAST	.00	-.01	-.27	.79
	CES-D	.02	.16	3.60	<.001
NDSS Drive	EMA WSWWS Crav. post-pre	.01	.09	1.97	.05
	MAST	.00	.06	1.33	.18
	CES-D	.00	.05	1.03	.30
WISDM Neg. Rein.	EMA WSWWS Neg. Affect post-pre	.01	.12	2.79	.01
	MAST	.01	.07	1.67	.10
	CES-D	.04	.19	4.35	<.001
FTND Item 1	# of cigs since last call pre-quit morning	.03	.17	3.76	<.001
	MAST	.00	.04	.88	.38
	CES-D	.00	.04	.80	.42
FTND Item 4	Cigs since yesterday pre-quit	.37	.62	18.17	<.001
	MAST	.00	.00	-.05	.96
	CES-D	.00	.02	.43	.67
NDSS Tol.	Cigs since yesterday pre-quit	.09	.31	7.71	<.001
	MAST	.00	.03	.64	.52
	CES-D	.00	-.03	-.64	.52
WISDM Tol.	Cigs since yesterday pre-quit	.10	.32	7.96	<.001
	MAST	.00	.04	.95	.34
	CES-D	.01	.07	1.67	.10

* Δr^2 denotes the change in r^2 between the model with only control variables to the model with the nicotine dependence measure added.

*** Post-pre indicates the difference between the post-quit scores and the pre-quit scores.