

# The construct of internalization: conceptualization, measurement, and prediction of smoking treatment outcome

G. SCOTT ACTON<sup>1\*</sup>, JEFFREY D. KUNZ<sup>2</sup>, MARK WILSON<sup>2</sup> AND SHARON M. HALL<sup>3</sup>

<sup>1</sup> *Rochester Institute of Technology, Rochester, NY, USA*; <sup>2</sup> *University of California, Berkeley, CA, USA*;

<sup>3</sup> *University of California, San Francisco, CA, USA*

## ABSTRACT

**Background.** Depression symptoms and diagnoses are associated with failure to quit smoking in most studies, but not all.

**Method.** A new measure of internalization (i.e. symptoms of depression or anxiety, or poor mood) was created to investigate whether internalization would predict smoking cessation in 549 smokers from three randomized clinical trials with inconsistent findings.

**Results.** Predicted item locations based on a map of the construct of internalization agreed with empirical locations based on item response theory. Internalization was highly correlated with neuroticism. Logistic regressions showed that internalization improved upon the predictions of other affect-related measures. High baseline internalization decreased abstinence from smoking at end of treatment and 3 months thereafter. History of major depression (single-episode or recurrent) failed to predict abstinence.

**Conclusions.** The broad, dimensional construct of internalization as conceptualized herein appears to be an important predictor of smoking cessation.

## INTRODUCTION

Internalization, a construct from the literature on child psychopathology (Zahn-Waxler *et al.* 2000) that has recently been applied to adult psychopathology (Krueger *et al.* 1998, 2003; Krueger, 1999; Vollebergh *et al.* 2001), is characterized by overcontrolled negative affect such as anxious-misery (including anxiety and depression) and fear. Partially analogous to the trait of neuroticism (Costa & McCrae, 1992), internalization refers to both the temporary and recurrent experience of negative affect. The importance of internalization as a construct derives in part from (a) the importance of establishing the continuity of childhood and

adult psychopathology; and (b) the well documented co-morbidity of anxiety and depression (Krueger & Finger, 2001), which may arise from cognitive (Mineka *et al.* 1998) and genetic (Kendler *et al.* 1992) affect-related vulnerabilities. By conceptualizing depression symptoms as only one aspect of the broader construct of internalization, by providing a new measure of internalization, and by using this measure to predict smoking treatment outcome, the authors hope to better understand the literature on the relation between depression symptoms or diagnoses and cigarette smoking.

## DEPRESSION AND SMOKING

The positive association between symptoms or diagnoses of depression and smoking has been shown repeatedly (Anda *et al.* 1990; Glassman *et al.* 1990; Glassman, 1993; Acton *et al.* 2001)

\* Address for correspondence: Dr G. Scott Acton, Department of Psychology, Rochester Institute of Technology, 92 Lomb Memorial Drive, Rochester, NY 14623-5604, USA.  
(Email: gsagss@rit.edu)

but not universally (Salive & Blazer, 1993). Depressive symptoms are generally associated with smoking. In a sample of Latino adults, current smokers had more depressive symptoms than former smokers or never smokers (Perez-Stable *et al.* 1990). Depressive symptoms were positively correlated with current smoking and negatively correlated with likelihood of quitting smoking (Anda *et al.* 1990). Depressive symptoms among 15- to 16-year-olds increased the likelihood of heavy smoking 9 years later (Kandel & Davies, 1986). Emotional distress in grade 10 increased the likelihood of smoking in grade 12, and smoking in grade 12 increased the likelihood of emotional distress in young adulthood (Orlando *et al.* 2001). In a contradiction of the above findings, however, older women (but not men) with depressive symptoms at baseline were nearly four times as likely to have quit smoking 3 years later (Salive & Blazer, 1993).

Diagnosed major depressive disorder (MDD) is also generally associated with smoking. In a population-based sample, persons with a history of MDD were over twice as likely to have a history of regular smoking than were persons with no MDD history (Glassman *et al.* 1990). In a random sample of young adults, those with moderate nicotine dependence had over four times the rates of MDD and anxiety disorders than those with no nicotine dependence (Breslau *et al.* 1991).

Negative affect and MDD are generally associated with smoking treatment outcome. Negative affect and history of MDD were positively correlated with smoking treatment failure (Glassman *et al.* 1988). Moreover, high negative affect following smoking cessation predicted treatment failure following behavioral counseling and a placebo (Covey *et al.* 1990). Smoking treatment success sometimes predicted MDD (Glassman *et al.* 2001) but sometimes did not (Tsoh *et al.* 2000).

Cognitive-behavioral treatment (CBT) is often effective for smoking (Muñoz *et al.* 1997; Patten *et al.* 1998). Despite some studies showing that CBT is differentially more effective for smokers with a history of MDD (Hall *et al.* 1994, 1998), it has become increasingly clear that history of MDD is not associated with smoking treatment outcome (Zelman *et al.* 1992; Hall *et al.* 1996; Brown *et al.* 2001). CBT may be differentially more effective for smokers with a

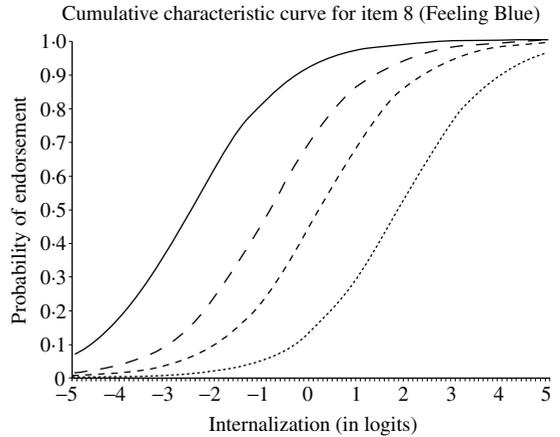


FIG. 1. Example of cumulative characteristic curves for a polytomous item. —, Threshold 1 (0–1); - - - -, threshold 2 (1–2); - · - ·, threshold 3 (2–3); ····, threshold 4 (3–4).

history of recurrent MDD (Brown *et al.* 2001; Haas *et al.* 2004).

An important unanswered question is whether the findings regarding MDD, depression symptoms, and smoking can be explained by a concept such as internalization, which includes other important phenomena, such as anxiety, which have been shown to relate to smoking. In the face of many unsuccessful prior attempts to quit, smokers may experience a combination of symptoms of depression and anxiety due to the interplay of felt helplessness and negative outcome expectancies (Mineka *et al.* 1998). Because internalization as a construct explicitly encompasses the co-morbidity of anxiety and depression (Krueger & Finger, 2001), as well as other aspects of poor mood (i.e. unpleasant mood states such as anger), it is expected that a dimensional measure of internalization will predict smoking treatment outcome better than the categorical construct of history of MDD.

## ITEM RESPONSE THEORY

Item response theory (IRT) assumes that any given item can be described as a function of an underlying latent dimension. The graphs of that function (see Fig. 1) are cumulative characteristic curves, which are nonlinear (logistic) regressions of probability of endorsement onto the latent dimension. The *x*-axis of a cumulative characteristic curve is the level of the latent dimension; the *y*-axis is a probability of endorsement

ranging from 0 to 1. In the case of polytomous items, separate curves for each threshold between response categories are generated. IRT, in which the focus is on the person in relation to the item, can be contrasted with classical test theory, in which the focus is on the standardization sample in relation to the scale (Embretson, 1996).

The point along the latent continuum at which the cumulative characteristic curve has a 0.5 probability of endorsement and provides the most information is called its *location*; location is the first item parameter, which is included in all IRT models. Inclusion of the location parameter, which is important for demonstrating the coherence of a construct, is one advantage of IRT over other methods, such as factor analysis. Analogous to the factor loading in factor analysis, *discrimination*, the second item parameter, is an index of how good the item is at measuring the latent dimension under consideration and is indicated by the maximum slope of the cumulative characteristic curve. Because a model with location as the only item parameter is simpler and has other desirable measurement properties, a one-parameter model that holds discrimination constant across all items is preferable whenever it provides an adequate fit to the data. Indeed, it may be desirable to construct tests that fit this one-parameter model [also called the Rasch model (Rasch, 1980)] whenever possible (Wright, 1977; Wilson, 2003).

## METHOD

### Analyses

Using IRT, we estimated the locations of a group of items on a latent dimension of internalization, assessed the relations of these items with a model of the construct of internalization, and assessed dimensionality, reliability, and validity. Using logistic regressions, we assessed the predictive ability of internalization, history of single-episode or recurrent MDD, and three treatments for smoking. The following analyses were performed.

(1) Item selection: the Internalization Scale IRT (ISIRT) was created by combining 169 items from four instruments into one item pool and selecting those items with raw scores showing Pearson correlations of  $r < -0.10$  with point-prevalence abstinence at 1 year following CBT.

Table 1. Sample sizes in each of three previous randomized clinical trials making up the samples in the present study

	Hall <i>et al.</i> (1994)	Hall <i>et al.</i> (1996)	Hall <i>et al.</i> (1998)
Inflated baseline	151	248	241
Original baseline	149	201	199
CBT	79	104	103
Nicotine gum	149	98	0
Nortriptyline	0	0	99

Participants not receiving cognitive-behavioral treatment (CBT) received a standard health-education group therapy. Participants in the first two studies (Hall *et al.* 1994, 1996) not receiving nicotine gum received a placebo; nicotine gum was not a treatment in the third study (Hall *et al.* 1998). Participants in the third study not receiving nortriptyline received a placebo; nortriptyline was not a treatment in the first two studies.

(2) The model: the correspondence between the internalization construct and the empirical locations of the ISIRT items was assessed.

(3) Dimensionality: one-dimensional and five-dimensional models of the ISIRT were compared using confirmatory IRT methods.

(4) Reliability: internal consistency and measurement information were assessed.

(5) Validity: item fit, concurrent validity, and cross-validity in different subsamples were assessed.

(6) Prediction: the predictive ability of internalization was compared with that of the four instruments from which items were selected; and internalization, history of single-episode or recurrent MDD, and treatment conditions (CBT, nicotine gum, and nortriptyline) were used to predict point-prevalence abstinence at end of treatment and 3 months following treatment.

### Participants

The current research was a reexamination of data from three previous randomized clinical trials (Hall *et al.* 1994, 1996, 1998). Participants were cigarette smokers recruited into a smoking cessation research program at the University of California, San Francisco (sample sizes appear in Table 1). The inflated sample included everyone in the original sample plus those who were assessed at baseline but did not enter the study owing to meeting an exclusion criterion. Those meeting criteria for current MDD (within the last 3–6 months) during the initial assessment were screened out because it was feared that, in

treatment, issues concerning current depression might overshadow smoking-related issues. Twenty-eight percent of participants were diagnosed as having a history of MDD by the Diagnostic Interview Schedule (Robins *et al.* 1981). Scores on the Beck Depression Inventory (Beck *et al.* 1961) were typically within the normal range for asymptomatic persons (mean = 7.16, s.d. = 6.14). Participants were predominantly Caucasian (89%); 53% were female. Average age was 40.2 years (range 20–65 years). Number of years smoked averaged 21.7, and age of beginning smoking averaged 14.8 years. Of the total sample, 283 participants successfully quit smoking by end of treatment and 166 maintained non-smoking to 3-month follow-up.

### Measures

The measures below were included because they are widely used to assess depression symptoms and related phenomena, and because they are the measures that were included in the baseline assessments of the earlier studies that were the source of the current data. The major interest of the present study was not so much the measures themselves, however, as the individual items drawn from these measures.

#### *Beck Depression Inventory (BDI)*

The BDI (Beck *et al.* 1961) is a widely used 21-item self-report questionnaire used to measure symptoms of depression that occurred during the past week. Item response options range from 0 to 3, with higher scores indicating more severe depression. The BDI shows acceptable test-retest reliability, internal consistency, and concurrent validity with major diagnostic systems.

#### *State-Trait Anxiety Inventory (STAI)*

The STAI (Spielberger *et al.* 1970) is a 40-item self-report instrument with two scales, one measuring how a person feels right now, that is, at this moment, and one measuring how a person generally feels. Response options range from 1 to 4. Anxiety is typically related to depression (Krueger & Finger, 2001).

#### *Profile of Mood States (POMS)*

The POMS (McNair *et al.* 1971) is a 65-item self-report instrument that assesses how a person has been feeling on the same day. The POMS provides a score on total mood disturbance as

well as on six subscales: Depression-Dejection, Tension-Anxiety, Anger-Irritability, Confusion, Fatigue, and Vigor. Response options range from 0 to 4. The subscales have shown excellent internal consistency reliability (0.89–0.95) and moderate test-retest reliability (0.65–0.74) (McNair *et al.* 1971). Clinical samples have been found to show higher POMS scores than non-clinical controls, and POMS scores are sensitive to various treatments, including both psychotherapy and pharmacotherapy (McNair *et al.* 1971). Negative mood states measured by the POMS have been found to predict relapse to smoking (Hall *et al.* 1983).

#### *Psychiatric Epidemiology Research Interview (PERI)*

A self-report adaptation of this interview (Dohrenwend *et al.* 1980) was used. Included were only the 43 items from scales related to internalization: Anxiety, Sadness, Psychophysiologic Symptoms, Poor Self-Esteem, Helplessness-Hopelessness, Dread, Confused Thinking, and Perceived Physical Health. Response options range from 0 to 4.

#### *Diagnostic Interview Schedule (DIS)*

The DIS (Robins *et al.* 1981) is a layperson-administered structured clinical interview. The present study used the computerized version of the DIS depression scale to assess MDD history.

#### *Revised Neuroticism, Extraversion, Openness Personality Inventory (NEO-PI-R)*

The NEO-PI-R (Costa & McCrae, 1992) is a widely used self-report measure of the five-factor model of personality. Only the 48 items dealing with neuroticism were administered. Neuroticism total score is made up of six subscales: Anger, Anxiety, Depression, Impulsivity, Self-Consciousness, and Vulnerability.

#### *California Psychological Inventory (CPI)*

The CPI (Gough, 1987) is a self-report measure of personality. Only the 19 items yielding a total score on Psychoneuroticism, a scale similar in content to neuroticism, were administered.

#### *Definition of successfully having quit smoking*

Three conditions needed to have been met in this study in order for a person to be considered

successfully to have quit smoking. First, the person must have provided a self-report of non-smoking in the past week. Second, breath samples must have shown carbon monoxide levels of  $\leq 10$  ppm. Third, urine samples must have shown cotinine levels of  $\leq 60$  ng/ml (Jarvis *et al.* 1987). If any of these conditions was not met, then the person was considered a treatment failure. The physiological measures provided an important check on a person's self-report. This stringent definition and its attendant conservative estimates of treatment success provide a trustworthy criterion for having attained abstinence.

## RESULTS

### Item selection

Items were selected by pooling 169 items from the BDI, STAI, POMS, and PERI, and correlating person raw scores on each item with 1-year point-prevalence abstinence from smoking among participants who had received CBT ( $n=286$ ). Item selection was limited to the CBT group in order to permit the very strong form of cross-validation assessed below. The 27 items with Pearson correlations arbitrarily set at  $r < -0.10$  with 1-year abstinence were retained for inclusion in the ISIRT.

### The model

#### Construct map

A model of the construct of internalization was created and tested against empirical findings. Internalization as a construct consists of a latent dimension along which persons and items can be located (Fig. 2). Predicted item locations (Table 2) were assigned based on a facet analysis that included two substantive hypotheses and one semantic convention; this facet analysis represented a theory or conjecture, not a procedure (Wilson, 2003). The first substantive hypothesis was that the affective domain can be divided into positive conditions (for example, joy) *versus* negative conditions (for example, anger). The second substantive hypothesis was that states are more extreme than traits (because states, being acute, are felt more intensely than chronic dispositions). Larsen & Diener (1987) postulated that *persons* differ in affect intensity, but the separation of person and

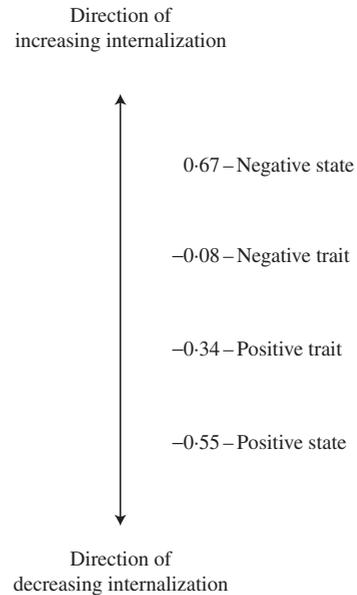


FIG. 2. Internalization construct map.

item parameters in IRT makes possible another conception of affect intensity, one specific to *items*. The model we have proposed is an item-explanatory model, which does not deal directly with either within-person differences in states over time or between-person differences in traits at the same time (for a discussion of the latter two, see Borsboom *et al.* 2003), but which could reasonably be applied to assess either kind of person differences. Rather, our model comprises a conjectured organization of the domain of items, such that items that assess relatively short-term characteristics (states) are hypothesized to be more intensely positive or negative than items that assess relatively long-term characteristics (traits). The semantic convention, that item responses can be divided into endorsement versus denial, was manifest in the procedure of reflecting positive items so that higher scores would indicate greater internalization.

Average item location within the resulting four types of item was taken as the predicted location for each item of that type. The fact that predicted item locations increased monotonically from low to high internalization provided a check on the validity of assignment to item types.

Table 2. Item analysis

Item type	ISIRT item no.	Original item no.	Item stem	Predicted item location	Observed item location ( $\beta$ )	Standard error of estimate
Negative state	1	BDI 1	Sadness	0.67	1.19	0.09
Negative state	2	BDI 2	Pessimism	0.67	1.20	0.09
Negative state	3	BDI 16	Insomnia	0.67	0.65	0.07
Negative trait	4	PERI 19	During the past year, how often have you feared crowds?	-0.08	0.38	0.06
Negative trait	5	PERI 21	During the past year, how often have you felt completely hopeless about everything?	-0.08	0.30	0.06
Negative trait	6	PERI 31	In general, how satisfied have you been with your body in the past year?	-0.08	-0.87	0.05
Negative trait	7	PERI 36	During the past year, how often has your appetite been poor?	-0.08	-0.33	0.06
Negative trait	8	PERI 41	During the past year, how often have you been bothered by feelings of sadness or depression - feeling blue?	-0.08	-0.32	0.05
Negative state	9	POMS 12	Peeved	0.67	0.24	0.06
Negative state	10	POMS 27	Restless	0.67	-0.14	0.06
Negative state	11	POMS 45	Ready to fight	0.67	0.42	0.06
Negative state	12	POMS 42	Desperate	0.67	1.09	0.08
Negative state	13	POMS 52	Deceived	0.67	0.38	0.06
Positive state	14	STATE 1R	Calm	-0.55	-0.20	0.07
Positive state	15	STATE 5R	At ease	-0.55	-0.37	0.06
Positive state	16	STATE 8R	Rested	-0.55	-0.79	0.06
Positive state	17	STATE 10R	Comfortable	-0.55	-0.33	0.06
Positive state	18	STATE 11R	Self-confident	-0.55	-0.08	0.07
Positive state	19	STATE 15R	Relaxed	-0.55	-0.50	0.06
Positive state	20	STATE 19R	Joyful	-0.55	-1.57	0.06
Positive state	21	STATE 20R	Pleasant (right now)	-0.55	-0.55	0.06
Positive trait	22	TRAIT 1R	Pleasant (generally)	-0.34	0.10	0.07
Positive trait	23	TRAIT 7R	'Calm, cool, and collected'	-0.34	-0.76	0.06
Negative trait	24	TRAIT 8	Difficulties piling up	-0.08	0.41	0.07
Negative trait	25	TRAIT 12	Lack self-confidence	-0.08	-0.16	0.07
Positive trait	26	TRAIT 13R	Secure	-0.34	-0.38	0.06
Negative state	27	BDI 9	Suicidal ideas	0.67	0.97*	N.A.

\* Parameter was constrained so that locations across items would be centered at 0.00 logits. Predicted item locations were based on the construct map (Fig. 2). Parameters were estimated using the inflated baseline sample from the first two studies (Hall et al. 1994, 1996) (n = 399). ISIRT, Internalization Scale Item Response Theory; BDI, Beck Depression Inventory; POMS, Profile of Mood States; STATE, State-Trait Anxiety Inventory State scale; R, reverse-scored; TRAIT, State-Trait Anxiety Inventory Trait scale.

Item locations

Observed item locations were estimated based on the partial credit model within the family of Rasch models (Masters, 1982), with number of response options differing among items. An IRT analysis of the 27 ISIRT items was performed on the inflated baseline sample from the first two studies (Hall et al. 1994, 1996) (n = 399) using the ACER CONQUEST computer software (Wu et al. 1998). This analysis resulted in location parameters (Table 2) for each item and for each threshold between item response categories (thresholds are discussed below). Item location was the average of the threshold locations.

The calibration resulted in item and person estimates that can be displayed in a location map (Fig. 3). Column 1 of the location map contains

ISIRT raw scores. Column 2 contains internalization scaled scores as measured in logits (or log-odds) - that is, the natural logarithm of the estimated probability of endorsing the item divided by the estimated probability of not endorsing the item, or  $\ln(p/q) = \ln(p/[1-p])$ , where  $p = e^{(\theta-\beta)}/[1+e^{(\theta-\beta)}]$ . Column 3 contains the distribution of person locations ( $\theta$ ). Column 4 contains the distribution of item locations ( $\beta$ ). The remaining columns contain the locations of item thresholds.

The average of the item distribution was set to 0 logits, allowing the person locations to vary freely. The person distribution was unimodal, ranging from 1 to -4 logits, and was centered at -1 logits (corresponding to a probability of endorsement of 0.27). Thus, these smokers were not high on internalization - a result to be



Table 3. *Translations of raw scores on the Internalization Scale Item Response Theory (ISIRT) into maximum-likelihood-estimated person locations ( $\theta$ ) on internalization, and the relation of internalization to abstinence from smoking*

Raw score on ISIRT (max. = 88)	Scaled score on internalization ( $\theta$ ) in logits	Probability of endorsing average item	Probability of abstinence at end of treatment	Probability of abstinence at 3-month follow-up	Standard error of measurement
86	3.83	0.98	0.19	0.11	0.72
84	3.11	0.96	0.22	0.13	0.51
82	2.69	0.94	0.25	0.14	0.42
80	2.38	0.92	0.26	0.15	0.37
78	2.13	0.89	0.28	0.16	0.33
76	1.93	0.87	0.29	0.17	0.31
74	1.75	0.85	0.30	0.17	0.29
72	1.60	0.83	0.31	0.18	0.27
70	1.46	0.81	0.32	0.18	0.26
68	1.33	0.79	0.33	0.19	0.25
66	1.21	0.77	0.33	0.19	0.24
64	1.09	0.75	0.34	0.19	0.23
62	0.99	0.73	0.35	0.20	0.23
60	0.88	0.71	0.36	0.20	0.23
58	0.78	0.69	0.36	0.21	0.22
56	0.69	0.67	0.37	0.21	0.22
54	0.59	0.64	0.37	0.21	0.22
52	0.50	0.62	0.38	0.22	0.22
50	0.40	0.60	0.39	0.22	0.22
48	0.31	0.58	0.39	0.22	0.22
46	0.21	0.55	0.40	0.23	0.22
44	0.12	0.53	0.41	0.23	0.22
42	0.02	0.50	0.41	0.24	0.22
40	-0.08	0.48	0.42	0.24	0.22
38	-0.18	0.46	0.43	0.24	0.23
36	-0.28	0.43	0.44	0.25	0.23
34	-0.39	0.40	0.44	0.25	0.24
32	-0.51	0.38	0.45	0.26	0.24
30	-0.63	0.35	0.46	0.26	0.25
28	-0.75	0.32	0.47	0.27	0.25
26	-0.88	0.29	0.48	0.28	0.26
24	-1.02	0.27	0.49	0.28	0.27
22	-1.17	0.24	0.50	0.29	0.28
20	-1.33	0.21	0.51	0.30	0.29
18	-1.51	0.18	0.52	0.31	0.30
16	-1.69	0.16	0.54	0.31	0.31
14	-1.90	0.13	0.55	0.32	0.33
12	-2.13	0.11	0.57	0.34	0.35
10	-2.39	0.08	0.59	0.35	0.38
8	-2.70	0.06	0.61	0.37	0.41
6	-3.08	0.04	0.63	0.39	0.46
4	-3.58	0.03	0.67	0.41	0.55
2	-4.38	0.01	0.72	0.46	0.74

An estimate of person location ( $\theta$  in logits can be found by summing response categories endorsed by the person (column 1) and reading over to the corresponding scaled score (column 2). All estimates except the probability of abstinence were based on inflated baseline data from the first two studies (Hall *et al.* 1994, 1996) ( $n = 399$ ). Estimates of the probability of abstinence were based on applying internalization scaled scores from this table to regression equations from the original baseline sample ( $n = 549$ ). Estimates of probability of endorsing an average item (i.e. an item located at  $\beta = 0.00$  logits) (column 3) were computed using the formula  $p = e^{(\theta - \beta)} / [1 + e^{(\theta - \beta)}]$ .

Scaled scores are in units of 1.00 logit, with origin at 0.00 logits. Estimates of probability of abstinence were based on the results of univariate logistic regressions: at end of treatment (column 4),  $p = e^{(-0.34 - 0.29\theta)} / [1 + e^{(-0.34 - 0.29\theta)}]$ ; at 3-month follow-up (column 5),  $p = e^{(-1.17 - 0.23\theta)} / [1 + e^{(-1.17 - 0.23\theta)}]$ .

and all responses below it for each ISIRT item. Although item locations were higher on internalization than were person locations, suggesting that higher levels of internalization were measured more reliably, the existence of multiple item thresholds within each item should

ensure reliable person measurement at lower levels of internalization as well.

*Predicted versus observed item locations*

Predicted locations of the ISIRT items based on the construct map (Fig. 2) were correlated with

observed locations displayed in the location map (Fig. 3) to establish the correspondence between the concept of internalization and its empirical realization (see also Table 2). The resulting correlation was high ( $r=0.76$ ), indicating that the ISIRT showed good agreement with theory.

### Dimensionality

We compared the one-dimensional partial credit model with a five-dimensional partial credit model to determine whether there were dimensional artifacts in our measure based on the fact that the items came from different instruments. Items were assigned to one of the five dimensions based on their source instrument: BDI, PERI, POMS, STATE, or TRAIT. The resulting model used 202 parameters (mean and variance for each latent dimension, 50 item locations, and 135 step parameters). To compare the fit of the two models, a formal statistical test was conducted by comparing the respective deviances. The deviance of the multi-dimensional model was 40841.24 with 202 degrees of freedom. The difference in deviance was 276.96 with 15 degrees of freedom. The difference in deviance was approximately  $\chi^2$ -distributed with the difference in degrees of freedom between the two models as the degrees of freedom. Because this value was statistically significant, the fit of the multi-dimensional model was deemed worse than the fit of the one-dimensional model. A comparison of the Akaike information criterion (AIC) (Akaike, 1987) yielded a similar result, with the one-dimensional model value of 40936.28 being smaller than the multi-dimensional value of 41243.24, indicating a better fit for the one-dimensional model.

### Reliability

#### Internal consistency

The internal-consistency reliability of persons was calculated using separation reliability, an internal-consistency metric for polytomous items that is interpreted in the same way as Cronbach's alpha (Wright & Masters, 1982). The separation reliability of persons was 0.89, indicating that the ISIRT measured these smokers well.

#### Measurement information

Under the Rasch model, standard error of measurement describes expected score fluctua-

Table 4. Correlations of the Internalization Scale Item Response Theory (ISIRT) with measures of related constructs

Scale	Correlation with internalization
NEO-PI-R neuroticism (total)	0.72
Depression	0.71
Vulnerability	0.61
Anxiety	0.59
Self-consciousness	0.58
Hostility	0.42
Impulsivity	0.29
CPI psychoneuroticism	0.54
MDD history	0.32

Correlations were based on inflated baseline data from the third study (Hall *et al.* 1998) ( $n=244$ ).

NEO-PI-R, Revised Neuroticism, Extraversion, Openness Personality Inventory; CPI, California Psychological Inventory; MDD, major depressive disorder.

tions in estimated person location due to error. Computed from expected *a posteriori* estimates using ACER Conquest (Wu *et al.* 1998), average standard error of measurement for these smokers was very low (mean = 0.09, s.d. = 0.05).

### Validity

#### Item fit

The item fit statistic is an indicator of how well the item fits the model. Calculated as the sum of squared residuals over persons for any one item (Wright & Masters, 1982), the index of item fit that we used was the weighted infit meansquare (Wu *et al.* 1998), which has an expected value of 1.0. It is possible for responses to an item to contradict the model by being either too orderly (thus denying the probabilistic nature of the model), as indicated by a weighted infit meansquare lower than 0.75, or by being too random, as indicated by a weighted infit meansquare greater than 1.33 (Wright & Masters, 1982). The only misfitting item was Item 7 (poor appetite), a case of unexpected randomness; the other ISIRT items fit quite well.

#### Concurrent validation

In order to help establish the validity of the ISIRT, maximum-likelihood-estimated scaled scores of persons on the ISIRT (see Table 3) in the inflated baseline sample in the third study (Hall *et al.* 1998) ( $n=244$ ) were correlated with raw scores on related measures. The results (Table 4) showed that internalization was highly

correlated with the Neuroticism total score and with the Depression, Vulnerability, Anxiety, and Self-Consciousness subscales of the NEO-PI-R; was moderately correlated with the Psychoneuroticism scale of the CPI; and was less correlated with the Impulsivity and Hostility subscales of the NEO-PI-R, which might be expected to relate more closely to externalization (Krueger *et al.* 2002; Acton, 2003) than to internalization. Theoretically and empirically, internalization is similar to but distinct from neuroticism, because internalization includes both states and traits. Internalization was distinct from MDD history assessed using the DIS. The convergence of internalization with theoretically related measures and the divergence with theoretically distinct measures underline the convergent and discriminant validity of the ISIRT.

#### Cross-validation

In order to cross-validate the ISIRT in a different group from the CBT group used in item selection, the interaction of CBT treatment condition with scaled-scores on the ISIRT in the original baseline sample ( $n=549$ ) was used to predict biologically verified point-prevalence abstinence. Multivariate logistic regressions were performed, with internalization, CBT, and the internalization-by-CBT interaction as predictors of abstinence from smoking. At neither end of treatment nor 3-month follow-up was the interaction statistically significant at  $p < 0.05$ . This result provided a very strong cross-validation of the ISIRT, not only in a different group but in a different kind of group (i.e. persons who did not receive CBT).

#### Prediction

The ability of internalization to predict abstinence from smoking was assessed by reanalyzing outcome data from the original three clinical trials (Hall *et al.* 1994, 1996, 1998) using scaled-scores on internalization in the original baseline sample ( $n=549$ ) to predict abstinence at the end of treatment and 3 months thereafter. The few cases with missing data for smoking abstinence were removed. The predictive ability of internalization was compared with that of the BDI, STAI, POMS, and PERI using univariate logistic regressions with each measure as the sole predictor of abstinence from smoking. The

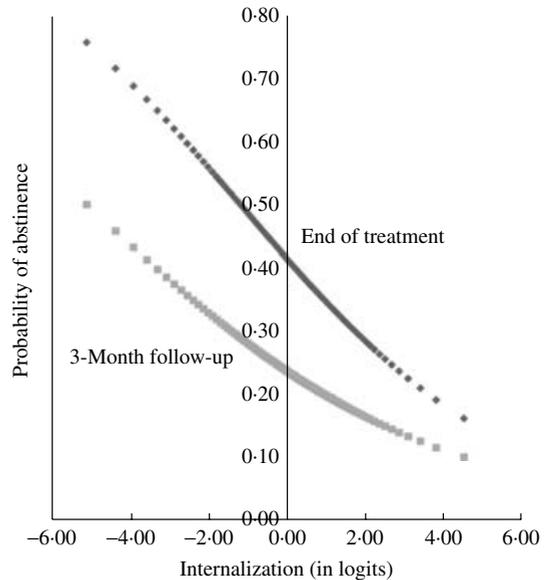


FIG. 4. Probability of abstinence from smoking as a function of baseline person locations ( $\theta$ ) on internalization.

predictive ability of internalization was not compared with that of neuroticism, because neuroticism data were not available from all three studies. The odds ratio (OR) was the estimate of effect size. An OR of 1.00 would indicate no effect, and a 95% Wald confidence interval (CI) that included 1.00 would indicate no statistically significant effect.

Internalization (OR 0.75, CI 0.62–0.91) was a good predictor of abstinence at the end of treatment (Table 3 and Fig. 4 show the probability of abstinence as a function of baseline scaled scores on internalization). The BDI (OR 0.96, CI 0.93–0.99), STAI State Anxiety (OR 0.98, CI 0.97–1.00), STAI Trait Anxiety (OR 0.98, CI 0.96–1.00), POMS Total Mood Disturbance (OR 0.96, CI 0.93–0.99), and PERI (OR 0.99, CI 0.98–1.00) were poor predictors. The fact that the confidence interval for internalization did not overlap that of the other predictors indicated that internalization was a better predictor than these measures at end of treatment.

Similarly, internalization (OR 0.80, CI 0.65–0.97) remained a good predictor of abstinence 3 months following the end of treatment. The BDI (OR 0.96, CI 0.93–1.00), STAI State Anxiety (OR 0.99, CI 0.97–1.00), STAI Trait

Table 5. *Multivariate logistic regressions for predictors of abstinence from smoking*

Parameter	df	Wald $\chi^2$	<i>p</i> value	Odds ratio
<b>End of treatment</b>				
Intercept	1	0.14	0.710	1.12
Internalization	1	5.31	0.021*	0.79
History of MDD	1	0.09	0.765	1.20
History of recurrent MDD	1	1.09	0.296	0.50
Cigarettes/week	1	12.59	0.000*	1.00
CBT	1	1.93	0.165	0.75
Nortriptyline	1	6.69	0.010*	2.16
Nicotine gum	1	16.47	0.000*	2.26
History of MDD $\times$ CBT	1	0.00	0.994	1.01
History of MDD $\times$ nortriptyline	1	0.02	0.902	0.89
History of recurrent MDD $\times$ CBT	1	1.68	0.195	2.90
History of recurrent MDD $\times$ nortriptyline	1	0.19	0.660	0.64
<b>Three months following end of treatment</b>				
Intercept	1	7.29	0.007*	0.40
Internalization	1	4.03	0.045*	0.81
History of MDD	1	0.40	0.528	1.47
History of recurrent MDD	1	1.67	0.197	0.41
Cigarettes/week	1	4.96	0.026*	1.00
CBT	1	0.50	0.480	1.17
Nortriptyline	1	7.35	0.007*	2.28
Nicotine gum	1	0.43	0.513	1.15
History of MDD $\times$ CBT	1	1.11	0.292	0.42
History of MDD $\times$ nortriptyline	1	0.00	0.973	1.03
History of recurrent MDD $\times$ CBT	1	2.78	0.096	4.49
History of recurrent MDD $\times$ nortriptyline	1	0.02	0.882	0.86

Scaled scores on internalization were estimated from raw scores on the Internalization Scale Item Response Theory (ISIRT) (see Table 3). Logistic regressions were based on original baseline data from all three studies ( $n=549$ ).

MDD, Major depressive disorder; CBT, cognitive-behavioral treatment.

\* Statistically significant at  $p < 0.05$ .

Anxiety (OR 0.98, CI 0.96–1.00), POMS Total Mood Disturbance (OR 0.99, CI 0.99–1.00), and PERI (OR 0.99, CI 0.98–1.00) remained poor predictors. Although the CI for internalization and the BDI overlapped, only the effect of internalization was statistically significant. The fact that the CI for internalization did not overlap that of the POMS or PERI indicated that internalization was a better predictor than these measures at 3-month follow-up.

Two multivariate logistic regressions of internalization scaled scores onto abstinence from smoking were performed, with history of single-episode or recurrent MDD, and treatment conditions – CBT, nicotine gum, and the tricyclic antidepressant nortriptyline (coded as active *versus* inactive) – also included as predictors of abstinence. The results indicated a significant treatment effect for nortriptyline and nicotine gum at end of treatment and for nortriptyline at 3-month follow-up (Table 5). MDD history failed to predict abstinence.

It was anticipated that smokers higher on internalization at baseline would be less successful

in quitting smoking and in maintaining non-smoking. The results (Tables 3 and 5, and Fig. 4) corroborated this expectation. The higher a smoker's initial internalization, the lower the probability of abstinence from smoking at end of treatment or 3 months thereafter.

## DISCUSSION

This study is consistent with previous studies regarding the prognosis for quitting smoking. It can now be more confidently asserted that high internalization is an important risk factor for failure to quit smoking. This may be because most adults who smoke have tried unsuccessfully to quit smoking, a negative event that may lead those with certain cognitive or genetic affect-related vulnerabilities to develop a combination of felt helplessness and negative outcome expectancies that lead to internalization and interfere with success in quitting smoking.

This study also contributes to an understanding of the phenomenology of internalization. Although the dimensionality of MDD has

previously been suggested by indirect and direct methods (Flett *et al.* 1997; Ruscio & Ruscio, 2000; see also De Boeck *et al.* 2005), and IRT analyses of depression measures have been performed (Orlando *et al.* 2000), an IRT-based measure of internalization that is both broader (for example, internalization instead of depression symptoms) and dimensional (for example, degree of internalization instead of an either/or construct) had yet to be created. It can now be more confidently asserted that internalization as a construct is a single dimension that is adequately reflected in the construct map (Fig. 2) and that the distribution of smokers on this construct approximates a normal distribution (Fig. 3).

In addition to the effect of internalization, the effects of history of single-episode and recurrent MDD and three treatments for smoking—CBT, nicotine gum, and nortriptyline—were examined. Nicotine gum and nortriptyline had an effect on abstinence, although the effect of nortriptyline lasted longer. The fact that internalization predicted abstinence, whereas MDD history did not, may suggest the relative importance of dimensional over categorical measurement in this domain.

A limitation of this study is that the sample of smokers was low on internalization by comparison with the ISIRT items. Thus, it is difficult to say whether the ISIRT could be used appropriately with persons high on internalization. Nevertheless, the measurement errors at low levels of internalization were low, reinforcing the conclusions drawn in this study. Moreover, the items themselves spanned the full range from low to high internalization.

A strength of this study is that internalization was a better predictor of abstinence from smoking at the end of treatment than the measures from which the ISIRT items were taken, and internalization continued to be a clearly better predictor than some measures (excluding the BDI) 3 months following the end of treatment. Internalization differs from the BDI and neuroticism in that theoretically and quantitatively it was shown to include both traits and states. The moderately low correlation with MDD history shows that internalization is not reducible to depression history. Instead, internalization is a broader construct than that assessed by any of the other measures, showing high correlations

with measures of several kinds of proclivity to experience poor mood.

This study was guided by the assumption that coherence (rational analysis or construct representation) as well as correspondence (predictive capacity or nomothetic span) is a desirable goal in measurement (Embretson, 1983). Typical measurement practice (for example, in construct validation) has focused primarily on correspondence. Inclusion of the location parameter based on IRT made it possible to test the coherence of an explicit map of the construct of internalization. Because the ISIRT showed both coherence (unidimensionality and a high correlation with the construct map) and correspondence (acceptable patterns of correlations with related measures and successful prediction of smoking cessation), the ISIRT is to be recommended as a valid measure of internalization. The ultimate goal of measurement, however, may be to make measures unnecessary by increasing understanding of the underlying phenomena. Our goal in presenting a new measure in this paper was to unite understanding of the phenomenology of internalization (coherence) with understanding of the prognosis for quitting smoking (correspondence). We believe our results indicate that goal at least partially to have been achieved.

## ACKNOWLEDGMENTS

Work on this article was supported in part by grants to G. Scott Acton from the National Institute on Drug Abuse (NIDA) (F32-DA14739) and the California Tobacco-Related Disease Research Program (10FT-0248) and by grants to Sharon M. Hall from NIDA (P50-DA09253, R01-DA02538, and T32-DA07250) and the National Cancer Institute (R01-CA71378). We are grateful to Paul De Boeck, Kevin L. Delucchi, and Janice Y. Tsoh for helpful comments on earlier versions of the article.

## DECLARATION OF INTEREST

None.

## REFERENCES

- Acton, G. S. (2003). Measurement of impulsivity in a hierarchical model of personality traits: implications for substance use. *Substance Use & Misuse* **38**, 67–83.

- Acton, G. S., Prochaska, J. J., Kaplan, A. S., Small, T. & Hall, S. M. (2001). Depression and stages of change for smoking in psychiatric outpatients. *Addictive Behaviors* **26**, 621–631.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika* **52**, 317–332.
- Anda, R. F., Williamson, D. F., Escobedo, L. G., Mast, E. E., Giovino, G. A. & Remington, P. L. (1990). Depression and the dynamics of smoking. *Journal of the American Medical Association* **264**, 1541–1549.
- Beck, A. T., Ward, C. H., Mendelsohn, M., Mock, J. & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.
- Borsboom, D., Mellenbergh, G. J. & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review* **110**, 203–219.
- Breslau, N., Kilbey, M. M. & Andreski, P. (1991). Nicotine dependence, major depression, and anxiety in young adults. *Archives of General Psychiatry* **48**, 1069–1074.
- Brown, R. A., Kahler, C. W., Niaura, R., Abrams, D. B., Sales, S. D., Ramsey, S. E., Goldstein, M. G., Burgess, E. S. & Miller, I. W. (2001). Cognitive-behavioral treatment for depression in smoking cessation. *Journal of Consulting and Clinical Psychology* **69**, 471–780.
- Costa Jr., P. T. & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and the NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources: Odessa, FL.
- Covey, L. S., Glassman, A. H. & Stetner, F. (1990). Depression and depressive symptoms in smoking cessation. *Comprehensive Psychiatry* **31**, 350–354.
- De Boeck, P., Wilson, M. & Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychological Review* **112**, 129–158.
- Dohrenwend, B. P., Shrout, P. E., Egri, G. & Mendelsohn, F. S. (1980). Nonspecific psychological distress and other dimensions of psychopathology: measures for use in the general population. *Archives of General Psychiatry* **37**, 1229–1236.
- Embretson, S. E. (1983). Construct validity: construct representation versus nomothetic span. *Psychological Bulletin* **93**, 179–197.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment* **8**, 341–349.
- Flett, G. L., Vredenberg, K. & Krames, L. (1997). The continuity of depression in clinical and nonclinical samples. *Psychological Bulletin* **121**, 395–416.
- Glassman, A. H. (1993). Cigarette smoking: implications for psychiatric illness. *American Journal of Psychiatry* **150**, 546–553.
- Glassman, A. H., Covey, L. S., Stetner, F. & Rivelli, S. (2001). Smoking cessation and the course of major depression: a follow-up study. *Lancet* **16**, 1929–1932.
- Glassman, A. H., Helzer, J. E., Covey, L. S., Cottler, L. B., Stetner, F., Tipp, J. E. & Johnson, J. (1990). Smoking, smoking cessation, and major depression. *Journal of the American Medical Association* **264**, 1546–1549.
- Glassman, A. H., Stetner, F., Walsh, B. T., Raizman, P. S., Fleiss, J. L., Cooper, T. B. & Covey, L. S. (1988). Heavy smokers, smoking cessation, and clonidine. *Journal of the American Medical Association* **259**, 2863–2866.
- Gough, H. G. (1987). *California Psychological Inventory Administrator's Guide*. Consulting Psychologists Press: Palo Alto, CA.
- Haas, A. L., Muñoz, R. F., Humfleet, G. L., Reus, V. I. & Hall, S. M. (2004). Influences of mood, depression history, and treatment modality on outcomes in smoking cessation. *Journal of Consulting and Clinical Psychology* **72**, 563–570.
- Hall, S. M., Bachman, J. S., Henderson, J., Barstow, R. & Jones, R. T. (1983). Smoking cessation in patients with cardiopulmonary disease: an initial study. *Addictive Behaviors* **8**, 33–42.
- Hall, S. M., Muñoz, R. F. & Reus, V. I. (1994). Cognitive-behavioral intervention increases abstinence rates for depressive-history smokers. *Journal of Consulting and Clinical Psychology* **62**, 141–146.
- Hall, S. M., Muñoz, R. F., Reus, V. I., Sees, K. L., Duncan, C., Humfleet, G. L. & Hartz, D. T. (1996). Mood management and nicotine gum in smoking treatment: a therapeutic contact and placebo-controlled study. *Journal of Consulting and Clinical Psychology* **64**, 1003–1009.
- Hall, S. M., Reus, V. I., Muñoz, R. F., Sees, K. L., Humfleet, G., Hartz, D. T., Frederick, S. & Triffleman, E. (1998). Nortriptyline and cognitive-behavioral therapy in the treatment of cigarette smoking. *Archives of General Psychiatry* **55**, 683–690.
- Jarvis, M. J., Tunstall-Pedoe, H., Feyerabend, C., Vesey, C. & Saloojee, Y. (1987). Comparison of tests used to distinguish smokers from non-smokers. *American Journal of Public Health* **77**, 1435–1438.
- Kandel, D. B. & Davies, M. (1986). Adult sequelae of adolescent depressive symptoms. *Archives of General Psychiatry* **43**, 255–262.
- Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. (1992). Major depression and generalized anxiety disorder: same genes, (partly) different environments? *Archives of General Psychiatry* **49**, 716–722.
- Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry* **56**, 921–926.
- Krueger, R. F., Caspi, A., Moffitt, T. E. & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM-III-R): a longitudinal-epidemiological study. *Journal of Abnormal Psychology* **107**, 216–227.
- Krueger, R. F. & Finger, M. S. (2001). Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. *Psychological Assessment* **13**, 140–151.
- Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G. & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum. *Journal of Abnormal Psychology* **111**, 411–424.
- Krueger, R. F., Chentsova-Dutton, Y. E., Markon, K. E., Goldberg, D. & Ormel, J. (2003). A cross-cultural study of the structure of comorbidity among common psychopathological syndromes in the general health care setting. *Journal of Abnormal Psychology* **112**, 437–447.
- Larsen, R. J. & Diener, E. (1987). Affect intensity as an individual difference characteristic: a review. *Journal of Research in Personality* **21**, 1–39.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika* **47**, 149–174.
- McNair, D. M., Lorr, M. & Droppelman, L. F. (1971). *Manual: Profile of Mood States*. EDITS: San Diego, CA.
- Mineka, S., Watson, D. & Clark, L. A. (1998). Comorbidity of anxiety and unipolar mood disorders. *Annual Review of Psychology* **49**, 377–412.
- Muñoz, R. F., Marin, B. V., Posner, S. F. & Perez-Stable, E. J. (1997). Mood management mail intervention increases abstinence rates for Spanish-speaking Latino smokers. *American Journal of Community Psychology* **25**, 325–343.
- Orlando, M., Ellickson, P. L. & Jinnett, K. (2001). The temporal relationship between emotional distress and cigarette smoking during adolescence and young adulthood. *Journal of Consulting and Clinical Psychology* **69**, 959–970.
- Orlando, M., Sherbourne, C. D. & Thissen, D. (2000). Summed-score linking using item response theory: application to depression measurement. *Psychological Assessment* **12**, 354–359.
- Patten, C. A., Martin, J. E., Myers, M. G., Calfas, K. J. & Williams, C. D. (1998). Effectiveness of cognitive-behavioral therapy for smokers with histories of alcohol dependence and depression. *Journal of Studies on Alcohol* **59**, 327–335.
- Perez-Stable, E. J., Marin, G., Marin, B. V. & Katz, M. H. (1990). Depressive symptoms and cigarette smoking among Latinos in San Francisco. *American Journal of Public Health* **80**, 1500–1502.
- Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. University of Chicago Press: Chicago (first published 1960).
- Robins, L., Helzer, J., Croughan, J. & Ratcliff, K. (1981). National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Archives of General Psychiatry* **38**, 381–389.

- Ruscio, J. & Ruscio, A. M.** (2000). Informing the continuity controversy: a taxometric analysis of depression. *Journal of Abnormal Psychology* **109**, 473–487.
- Salive, M. E. & Blazer, D. G.** (1993). Depression and smoking cessation in older adults: a longitudinal study. *Journal of the American Geriatrics Society* **41**, 1313–1316.
- Spielberger, C. D., Gorsuch, R. L. & Lushene, R. E.** (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press: Palo Alto, CA.
- Tsoh, J. Y., Humfleet, G. L., Muñoz, R. F., Reus, V. L., Hartz, D. T. & Hall, S. M.** (2000). Development of major depression after treatment for smoking cessation. *American Journal of Psychiatry* **157**, 368–374.
- Vollebergh, W. A. M., Iedema, J., Bijl, R. V., de Graaf, R., Smit, F. & Ormel, J.** (2001). The structure and stability of common mental disorders: the NEMESIS study. *Archives of General Psychiatry* **58**, 597–603.
- Wilson, M.** (2003). On choosing a model for measuring. *Methods of Psychological Research* **8**, 1–22.
- Wright, B. D.** (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement* **14**, 97–116.
- Wright, B. D. & Masters, G. N.** (1982). *Rating Scale Analysis: Rasch Measurement*. MESA Press: Chicago.
- Wu, M. L., Adams, R. J. & Wilson, M.** (1998). *ACER CONQUEST: Generalized Item Response Modelling Software*. Australian Council for Educational Research: Melbourne, Australia.
- Zahn-Waxler, C., Klimes-Dougan, B. & Slattery, M. J.** (2000). Internalizing problems of childhood and adolescence: prospects, pitfalls, and progress in understanding the development of anxiety and depression. *Development and Psychopathology* **12**, 443–466.
- Zelman, D. C., Brandon, T. H., Jorenby, D. E. & Baker, T. B.** (1992). Measures of affect and nicotine dependence predict differential response to smoking cessation treatments. *Journal of Consulting and Clinical Psychology* **60**, 943–952.