



RELIABILITY AND VALIDITY OF THE YALE-BROWN OBSESSIVE-COMPULSIVE SCALE

SHEILA R. WOODY,^{1*} GAIL STEKETEE² and DIANNE L. CHAMBLESS³

¹Department of Psychology, Yale University, P.O. Box 208205, New Haven, CT 06520, U.S.A.

²School of Social Work, Boston University, 264 Bay State Rd, Boston, MA 02215, U.S.A. and

³Department of Psychology, American University, 4400 Massachusetts Ave, NW, Washington, DC 20016, U.S.A.

(Received 21 September 1994)

Summary—The reliability and validity of the Yale-Brown Obsessive-Compulsive Scale were examined according to a multi-trait multi-method approach in a sample of 54 outpatients with obsessive-compulsive disorder (OCD). Internal consistency was acceptable but was improved by deletion of items concerning resistance to obsessions and compulsions. Inter-rater reliability was excellent, but test-retest reliability over an average interval of 48.5 days was lower than desirable. The YBOCS demonstrated good convergent validity with most other measures of OCD, but divergent validity *vis à vis* depression was poor. Analyses of new items assessing avoidance and the duration of obsession-free and compulsion-free intervals indicated that only the avoidance rating added meaningfully to the full scale score. In future research the authors recommend deletion of the resistance items and inclusion of the avoidance item to yield a revised 9-item YBOCS total score.

INTRODUCTION

Research on the psychopathology and treatment of obsessive-compulsive disorder (OCD) is a highly active area for both behavioral and pharmacological investigators (see Steketee, 1993). Perhaps because of the unusually wide variability in the content of symptoms associated with obsessive-compulsive disorder, many different instruments have been developed to measure the severity of these symptoms for research purposes. Few instruments have been adopted consistently across investigations, either because of criticisms of their psychometric properties or because of concerns about their clinical accuracy. The present study examines the psychometric properties of the Yale-Brown Obsessive Compulsive Scale (YBOCS: Goodman, Price, Rasmussen, Mazure, Fleischmann, Hill, Heninger & Charney, 1989a; Goodman, Price, Rasmussen, Mazure, Delgado, Heninger & Charney, 1989b). The YBOCS has been widely adopted to assess outcome, mainly in medication trials, but its validity and reliability have yet to be amply documented.

The YBOCS, a semi-structured interview, was introduced by Goodman and colleagues to measure obsessive compulsive symptoms in a content-free format (Goodman *et al.*, 1989a, b). That is, the scale ratings do not depend on specific types of symptoms (e.g. washing, checking, counting), but are instead based on aspects of those symptoms as reported by the patient during the interview (e.g. duration, interference, degree of resistance). The scale is divided into two parts of five questions each, the Obsessions subscale and the Compulsions subscale. On each subscale five aspects of pathology are rated on scales ranging from 0 (no symptoms) to 4 (extreme symptoms): (1) time spent, (2) degree of interference, (3) distress, (4) resistance, and (5) perceived control over the symptom. Unlike some other OCD measures, the YBOCS assigns lower scores to greater resistance as an indicator of health. Scores obtained from the subscales are summed to yield a YBOCS Total score.

The findings from several investigations of the psychometric properties of the YBOCS are summarized in Table 1. The results of tests of internal consistency, inter-rater reliability, and test-retest reliability in clinical and nonclinical samples have been excellent (Frost, Steketee, Krause & Trepanier, 1994; Goodman *et al.*, 1989a; Kim, Dysken & Kuskowski, 1990, 1992). However,

*Author for correspondence.

findings on test-retest reliability in particular require replication in that these were only reported by one research group (Kim *et al.*, 1990; 1992) who used the same rater and a test-retest interval of only 1 week. With such a short time between assessments, the interviewer's ratings may well have been colored by recollections of the recent prior interview, thus inflating reliability estimates.

Examining convergent validity of the YBOCS presents some difficulties, since there is no agreed-upon gold standard measure for comparison. Some researchers have examined the YBOCS in relation to measures commonly used to assess symptoms of OCD. Gathering data over a series of three different pharmacological treatment studies, Goodman *et al.* (1989b) found mixed evidence of validity. The YBOCS showed good convergence with two of three commonly used OCD measures, the NIMH-OC scale and the Clinical Global Impression-OC score (both of which are assessor-rated scales), but not with the self-report Maudsley Obsessional Compulsive Inventory (MOCI; Hodgson & Rachman, 1977). The latter measure itself suffered from poor correlations with other OCD measures in those studies, but the sample size was too small (*n*s of 10 and 13) for confidence in these findings.

Kim *et al.* (1990) also reported mixed results for YBOCS convergent validity. Although strongly related to physician and patient global ratings of improvement, the YBOCS did not correlate well with the Leyton Obsessional Inventory (LOI). Note, however, that some researchers have expressed concerns about the validity of the LOI for assessment of OCD (e.g. Emmelkamp, 1988a). The lack of strong convergent validity with some measures of OCD symptoms in clinical samples may merely reflect greater accuracy of some instruments in detecting these symptoms, especially idiosyncratic obsessions and compulsions. When validity of the YBOCS was studied in 43 non-clinical Ss, findings indicated significant and moderately strong correlations of the YBOCS with three measures of OC symptoms: the MOCI, the revised Compulsive Activity Checklist, and the Obsessive Thoughts Questionnaire (Frost *et al.*, 1994).

Table 1. Previous reports on psychometric properties of the Yale-Brown obsessive compulsive scale

	Goodman <i>et al.</i> (1989a)	Goodman <i>et al.</i> (1989b)	Kim <i>et al.</i> (1990)	Kim <i>et al.</i> (1992)	Frost <i>et al.</i> (1994)
Sample	40 OCD patients	81 OCD patients from 3 separate studies	23 nondepressed OCD patients	28 nondepressed OCD patients	43 non-clinical students
<i>Reliability</i>					
Alpha	0.88 to 0.91	—	—	—	0.88
Item-total correlation	$r = 0.36$ (item 4) to 0.77	—	—	—	$r = 0.43$ to 0.75
Inter-rater ICCs	YBOCS Total $r = 0.98$ Obsessions $r = 0.97$ Compulsions $r = 0.96$ items 0.86–0.97	—	—	—	94% agreement ($n = 10$ cases)
Test-retest ICCs	—	—	$r = 0.90$	$r = 0.97$	—
<i>Convergent validity</i>					
Physical global	—	$r = 0.74$	$\rho = -0.73^{*\dagger}$	$\rho = 0.82^*$	—
Patient global	—	—	$\rho = 0.54^{*\dagger}$	$\rho = 0.72^*$	—
NIMH-OC	—	$r = 0.67$ ($n = 20$)	—	$\rho = 0.80^*$ baseline, $\rho = 0.77$	—
MOCI	—	$r = 0.53$ ($n = 66$)	—	—	$r = 0.58$
LOI	—	—	Total $\rho = 0.3$ Symptom $\rho = 0.38$	—	—
SCL-90-R OC	—	—	—	$\rho = 0.41^*$ baseline, $\rho = 0.17$	—
CAC-R	—	—	—	—	$r = 0.55$
OTC	—	—	—	—	$r = 0.64$
<i>Divergent validity</i>					
Hamilton-Depression	—	$r = 0.60$ ($n = 80$)	—	—	—
Hamilton-Anxiety	—	$r = 0.47$ ($n = 79$)	—	—	—

Note: Psychometric properties are for the total (10 item) score, unless specified otherwise.

*These correlations were performed on simple change scores (termination minus baseline). Baseline scores are noted if available.

†Negative correlations are not explained in the article.

ICC = Intra-class correlations.

CAC-R = Compulsive Activity Checklist—Revised.

LOI = Leyton Obsessional Inventory.

MOCI = Maudsley Obsessional Compulsive Inventory.

OTC = Obsessive thoughts Questionnaire.

SCL-90-R = Symptom Checklist 90—Revised.

Only one test of divergent validity for the YBOCS has been published (Goodman *et al.*, 1989b). The results, based on one inpatient and two outpatient samples, were not impressive due to strong positive correlations between the YBOCS and the Hamilton rating scales for anxiety and for depression. In the small inpatient sample ($n = 16$) which had the broadest range of depression and anxiety, the YBOCS showed a stronger relationship with the Hamilton mood scales than with OCD measures (see Table 1). Since OCD is classified as an anxiety disorder and is commonly comorbid with depression, it is likely that any measure of OCD symptoms will also be substantially associated with both anxious and depressive symptoms. None the less, stronger association with other measures of obsessive and compulsive symptoms would be expected.

Recent evidence of the construct validity of the YBOCS comes from a factor analysis conducted by Danyko, McKay and Neziroglu (1993). As would be predicted, these authors found a two-factor solution to the responses of 83 patients with OCD. Distinct, relatively uncorrelated factors for obsessions and compulsions emerged, with only the Resistance item from the Obsession subscale failing to load significantly on either factor.

Thus, data gathered from several independent sources indicate that the YBOCS total score has good reliability and convergent, but not divergent, validity. Several additional questions remain. There have been no published psychometric data for the recently added Items 1a and 6a (obsession-free and compulsion-free intervals) or for Items 11 through 19 of the YBOCS, which were designated as investigational at the time of the original studies by Goodman and colleagues. Of particular interest among these is Item 12, Avoidance. Avoidance of situations that provoke distress or rituals is a marked feature of many clients with OCD and is contained in most other assessment instruments, but has not been included among the main items of the YBOCS.

In addition, although the total score has been subjected to close scrutiny, the two subscales have not, with the exception of the Danyko *et al.* (1993) factor analysis. Further evidence of the subscales' psychometric soundness would increase confidence in their usefulness for comparing the effects of various treatments on obsessions and compulsions. Also sorely needed is some indication of criterion-related validity of the YBOCS, for example with observed obsessive-compulsive behaviors in a challenge situation such as a behavioral avoidance test.

METHOD

Participants

Fifty-four outpatients with a primary diagnosis of OCD participated in this study. All were clients scheduled to begin behavioral treatment at McLean Hospital in Belmont, Mass. or the American University Agoraphobia and Anxiety Program in Washington, DC. Potential clients were excluded if there was evidence of current or past psychosis, current alcohol or substance dependence, or organicity. Only clients engaged in overt rituals for at least 1 h per day were admitted. Of the participants, 65% were female; 92% were Caucasian, 4% African-American, and 4% from other ethnic backgrounds. Their mean age was 33.2 years ($SD = 9.8$).

Measures

Interview measures. Primary diagnosis was determined by an experienced clinician (D.L.C. or G.S.) and corroborated by trained interviewers using the Structured Clinical Interview for *DSM-III-R* (SCID-P; Spitzer, Williams, Gibbon & First, 1990a). Interviewers were doctoral students in clinical or counseling psychology or social work. Interviewers observed a videotape of a master SCID interviewer and live SCIDs administered by the two senior investigators (D.L.C. and G.S.) or other experienced interviewers, and were supervised during their early administrations of the SCID by an experienced interviewer via live or taped observation until they were judged proficient.

In conjunction with the SCID interviews, clients were assessed using the YBOC Symptom Checklist to identify their specific obsessions and compulsions, followed by administration of the 19-item YBOCS interview. YBOCS training was conducted by the second author (G.S.) following a training workshop with Dr Goodman and additional correspondence with him regarding rating questions. Training procedures paralleled those for the SCID.

SCID and YBOCS interviews were audiotaped and 22 SCID tapes and 30 YBOCS tapes were coded by a second blind rater to study cross-site reliability. Kappa for the SCID diagnosis of OCD was 1.0.

Self-report questionnaires. The Maudsley Obsessional–Compulsive Inventory (MOCI; Hodgson & Rachman, 1977), a 30-item true–false questionnaire frequently used in treatment outcome research, was included as a standardized self-report measure of OCD symptoms. This measure is internally consistent and has satisfactory test–retest reliability. It discriminates OCD patients well from those with other anxiety disorders and anorexia, but not from those with depression (see review by Emmelkamp, 1988b).

The self-report Symptom Checklist 90-R anxiety and depression subscales were used for measurement of discriminant validity. Both subscales have shown high internal and test–retest reliability (Derogatis, 1977). Derogatis reports impressive evidence of convergent and criterion-related validity for the depression subscale, but provides less specific data for the anxiety subscale.

Behavioral avoidance tests. BATs were administered by project therapists before the exposure and response prevention phase of the treatment began. Three situations that were difficult or impossible for the clients to do without significant anxiety or rituals were selected and attempted (e.g. touching a so-called contaminated object without washing; driving on a busy street without checking). These tests were conducted *in vivo*, and anxiety levels (0–100) and degree of avoidance or interference from rituals were recorded for each situation (0 = no avoidance/rituals, 1 = partial avoidance/rituals, 2 = unable to do task). Scores were then averaged across tasks. Freund (1986) has reported good convergent validity of a similar BAT for OCD with self-reported avoidance, but the test–retest reliability of BATs with OCD patients is unknown.

Target symptoms. Obsessions and compulsions were identified by the therapist during the pretreatment assessment. These reflected the idiosyncratic major concerns of each client. Clients rated their fear and/or avoidance (Fear/Avoid) of up to three personally relevant situations (e.g. contamination from HIV) and the frequency and/or duration of up to two personally relevant rituals (e.g. handwashing) on scales ranging from 0 (no symptoms) to 8 (severe symptoms). These scales are widely used in behavioral research on OCD. Freund (1986) has reported good inter-rater and test–retest reliability for two of three target rating scales: Ratings of avoidance and rituals proved reliable, but fear was less so.

Procedure

Prospective clients attended an intake interview with one of the senior investigators. Those who met preliminary inclusion criteria for one of the protocols currently underway at the particular site were referred for a SCID. Clients with a SCID diagnosis of OCD were also administered the YBOCS at this time. All clients provided informed consent before the interviews. To collect data on test–retest reliability for the YBOCS, the intake interviewer also administered the YBOCS to 24 clients with OCD. The second interviewer was uninformed as to the initial ratings. The test–retest interval ranged from 10–103 days ($M = 48.5$).

During the first two sessions which were devoted to treatment planning, the therapists determined the items to be used for the target ratings and for the BAT. OCD clients then completed these additional assessment procedures before beginning 16 sessions of exposure and response prevention followed by 4 maintenance sessions. Treatment was conducted over a 4-month period. Following treatment, clients again completed all measures of OC symptoms. At the time of this report, 33 of the OCD clients had completed treatment and provided YBOCS post-test data.

RESULTS

Of major interest to this investigation were the properties of the YBOCS subscale scores, the total score, and the investigational Avoidance item. All proved to be normally distributed. Not surprisingly, total scores ranged in the upper two-thirds of the possible range for this treatment-seeking sample, but avoidance scores were well distributed across the 0–4 possible range of scores. We also examined whether the two “free interval” items (1a and 6a) were a useful addition to scale.

Table 2. Yale-Brown obsessive compulsive scale: internal consistency

	Item-remainder correlation ($n = 51$) (Subscale)	Item-remainder correlation ($n = 51$) (YBOCS Total)
<i>Obsessions Subscale</i>	alpha = 0.77	
1. Time spent	0.62	0.48
2. Interference	0.53	0.53
3. Distress	0.63	0.51
4. Resistance	0.34	0.30
5. Control	0.60	0.51
<i>Compulsions subscale</i>	alpha = 0.51	
6. Time spent	0.27	0.34
7. Interference	0.43	0.38
8. Distress	0.21	0.38
9. Resistance	0.09	-0.10
10. Control	0.45	0.21
<i>YBOCS total</i>	alpha = 0.69 ^a	

^aAlpha for a 9-item YBOCS excluding both Resistance items and including Avoidance (item 12) is 0.78.

Reliability

Internal consistency. Item-total correlations and Cronbach's alphas were calculated for various permutations of the YBOCS and are presented in Table 2. Internal consistency for the YBOCS Total score (Items 1-10) was acceptable (0.69). However, Item 9 (Resistance to Compulsions) was not correlated with the Total score ($r = -0.10$). Without this item, alpha for the Total scale increased to 0.74.

When the subscales were examined separately, alpha was acceptable for Obsessions but low for Compulsions. Again, Item 9 failed to correlate with the Compulsion subscale score. When we dropped Resistance from the Compulsion subscale, alpha increased from 0.51 to 0.58. In addition, Items 6 (Time Spent) and 8 (Distress) had item-total correlations below 0.30, indicating that the Compulsion subscale is considerably less consistent than the Obsession subscale.

Later versions of the YBOCS contain two additional experimental items, 1a "Obsession-free Interval" and 6a "Compulsion-free Interval", with higher scores indicating shorter intervals between symptom occurrence. No data have been published regarding their psychometric properties. Including these items in calculations of internal consistency yielded alphas of 0.70 for YBOCS Total, 0.78 for Obsessions subscale, and 0.53 for Compulsions subscale. These results are virtually identical to alphas calculated without these items (See Table 2). The obsession-free interval was highly correlated with Obsessions Total score ($r = 0.80$), but the compulsion-free interval was not strongly related to the Compulsions Total score ($r = 0.24$). Symptom-free intervals on both scales were strongly related to the "Time spent" on obsessions ($r = 0.75$, $P < 0.0001$, $n = 29$) and compulsions ($r = 0.68$, $P < 0.0001$, $n = 29$), respectively. These data suggest that these two experimental items are largely redundant with existing items and could be eliminated. Accordingly, they will not be considered further in this paper.

On the other hand, Item 12, Avoidance, appears to hold promise for adding meaningfully to YBOCS Total. This item proved to be strongly associated with YBOCS Total ($r = 0.51$, $P = 0.0001$) and YBOCS Compulsions ($r = 0.63$, $P = 0.0001$), although not with Obsessions ($r = 0.24$, $P > 0.09$). When an 11-item YBOCS Total was constructed, including Avoidance as an additional item, Cronbach's alpha rose from 0.69 to 0.73. Finally, because we found the Resistance to Compulsions item to be problematic, and Danyko *et al.* (1993) reported problems with the Resistance to Obsessions item, we dropped these two items, added Avoidance, and recalculated alpha for YBOCS Total. Alpha rose to 0.78.

Inter-rater reliability. Reliability was calculated with intraclass correlation coefficients which are presented in Table 3. Reliability was excellent for YBOCS Total and the two subscales and good for the single item Avoidance rating.

Test-retest reliability. Evidence of test-retest reliability was examined using intraclass correlations (see Table 3). The scores for YBOCS Total and the subscales proved less reliable than desirable, although still acceptable for research purposes (see Fleiss, 1986),* especially given the

*Standards for reliability for clinical purposes, when decisions are to be made about an individual client on the basis of an assessment instrument, need to be particularly stringent.

Table 3. Inter-rater and test-retest reliability for YBOCS items and scale scores

	Inter-rater ICCs (<i>n</i> = 30)	Test-retest ICCs (<i>n</i> = 24)
<i>Obsessions subscale</i>	0.94	0.64
1. Time spent	0.91	0.55
2. Interference	0.84	0.54
3. Distress	0.85	0.62
4. Resistance	0.81	-0.09
5. Control	0.90	0.56
<i>Compulsions subscale</i>	0.89	0.56
6. Time spent	0.89	0.43
7. Interference	0.76	0.60
8. Distress	0.80	0.37
9. Resistance	0.78	0.55
10. Control	0.82	0.42
<i>YBOCS total (1-10)</i>	0.93	0.61
12. Avoidance	0.70	0.51

Note: ICC = Intra-class correlation.

change in raters and lengthy duration of the intertest interval. Most individual items showed moderate correlations, but reliability for some (in particular Resistance to Obsessions) fell in the poor range. Given the modest reliability coefficients, it is important to determine whether scores change in a systematic fashion across a period with no treatment. For example, the initial administration might have a pretest sensitization effect, leading later scores to be notably higher. Repeated measures *t* tests demonstrated that YBOCS Total ($t = 0.75$, $P = 0.46$), Obsessions ($t = 0.88$, $P = 0.39$), and Compulsions ($t = 0.36$, $P = 0.72$) did not systematically increase or decrease over time.

Validity

Construct validity. YBOCS Total was not significantly influenced by the participants' sex, age, or socioeconomic status (all P s > 0.27). Sex and SES also had no significant effects on either of the subscale scores. However, age was significantly and positively correlated with Compulsions ($r = 0.39$, $P < 0.005$), such that older clients reported more severe rituals. Since pretest Obsession subscale scores tended to correlate negatively with age ($r = -0.24$, $P < 0.09$), the significant positive relationship with Compulsions at pretest only may be a random fluctuation.

To examine convergent validity, we calculated correlations of the YBOCS subscale and Total scores with other standard measures of OCD pathology taken at pretest, the MOCI, Target Symptom Ratings, and the Behavioral Avoidance Test (see Table 4). The Target Symptom Ratings for Fear/Avoidance and Rituals were composited for comparison with the YBOCS Total score, to make the comparison more conceptually homogenous. The YBOCS Total score showed modest to moderate correlations with all self-report and behavioral measures of OCD, although the correlation with the Target Ratings composite fell short of significance ($P = 0.09$).

At pretreatment, the YBOCS Obsessions subscale was moderately correlated with the MOCI Total score and with BAT anxiety but, unexpectedly, not with Target Ratings of Fear/Avoidance

Table 4. Pearson correlations between the YBOCS and measures of OC symptoms and moodstate at pre- and post-treatment

	Pretest				Posttest			
	<i>n</i>	Obs.	Com.	Total	<i>n</i>	Obs.	Com.	Total
<i>MOCI</i>	50	0.33*	0.35*	0.43**	30	0.53**	0.48*	0.55**
<i>SCL-90</i>								
Anxiety	46	0.33*	0.00	0.23	28	0.43*	0.24	0.37
Depression	46	0.33*	0.35*	0.42**	28	0.48*	0.46*	0.51*
<i>Target ratings</i>								
Fear/Avoid	44	0.18	—	—	29	0.31	—	—
Rituals	44	—	0.30*	—	29	—	0.72***	—
Composite	44	—	—	0.26	29	—	—	0.64***
<i>BAT</i>								
Avg. SUDS	34	0.36*	0.30	0.43*	27	0.35	0.06	0.23
Avoidance/ Rituals	42	0.27	0.32*	0.38*	27	0.32	0.36	0.37

*** $P < 0.0005$; ** $P < 0.005$; * $P < 0.05$ (all two-tailed).

(Table 4). Item 12, Avoidance, showed good evidence of convergent validity in that it was significantly related to Target Ratings Fear/Avoidance ($r = 0.31$, $P < 0.05$) and to overall Avoidance/Rituals on the BAT ($r = 0.43$, $P < 0.004$). The Compulsions subscale was moderately correlated with the MOCI and with overall Avoidance/Rituals on the BAT (Table 4).

The low correlations of the YBOCS scales with target ratings are particularly troublesome. However, these may result from restriction of range on the target ratings before treatment when almost all clients rated themselves as quite severely disturbed. Accordingly, we also examined the correlations of the YBOCS with these measures after treatment. At this point target ratings (as well as other instruments) had a considerably wider range, reflecting varying degrees of improvement. Consistent with the restriction of range hypothesis, the magnitude of the correlations between the YBOCS and target ratings increased. Indeed with the increased range on all instruments, almost all correlations among OC measures rose after treatment* (Table 4).

To study divergent validity of the YBOCS, Pearson correlation coefficients were calculated for the Total score in relation to SCL-90-R subscale measures of depression and general anxiety. These results are also presented in Table 4. The YBOCS Total score was moderately associated with depression at both pre- and post-test. Indeed, at both time periods, the YBOCS was as strongly related to the depression subscale of the SCL-90-R as to other measures of OCD symptoms. These findings indicate poor divergent validity, particularly at pretest: The YBOCS did not distinguish well between constructs of OC pathology and depression.

To further explore the relationship of the YBOCS to depression, we conducted additional analyses. When clients with diagnoses of Major Depression or Dysthymia were removed from the sample, pretest correlations of SCL-90-R depression with YBOCS variables were essentially unchanged. Next, scores of clients with concurrent diagnoses of depression or dysthymia were contrasted with those of clients without comorbid depression. Depressed and nondepressed clients' YBOCS scores did not differ [$t(49) = -1.47$, $P < 0.15$]. These findings suggest that comorbid depressive disorders have little impact on YBOCS scores and, therefore, that the strong relationship of the YBOCS to measures of depression may reflect increased demoralization in the more symptomatic OC clients.

With regard to divergent validity *vis à vis* anxiety, findings were positive for the YBOCS Total and Compulsion Scales. On the whole, correlations of these variables with other measures of OCD are substantially stronger than with anxiety. However, the Obsessions Scale proved to be as correlated with anxiety as with OCD measures. Although some overlap would be expected, the relative size of the correlations does not support divergent validity for Obsessions.

DISCUSSION

In general the YBOCS 10-item scale performed reasonably well psychometrically, although results from this study were less positive than those of its originators (Goodman *et al.*, 1989a, b). Items spanned the range of the scale, allowing for confidence in the appropriateness of items for assessing the substantial variability of OCD symptom severity, and allowing for confidence in later analyses using correlations.

Findings for internal consistency ($\alpha = 0.69$) were acceptable, but well below the 0.89 figure given by Goodman *et al.* (1989a) for OCD *Ss* and the 0.88 figure of Frost *et al.* (1994) for *Ss* with subclinical OCD. Item-total correlations indicated that Item 9, Resistance to Compulsions, was very poorly related to the overall scale and, when included in the alpha for the Compulsions subscale, it reduced internal consistency of that scale to below acceptable levels (0.51). Further indication that items concerning resistance may prove problematic comes from Danyko *et al.*'s (1993) findings that the Resistance to Obsessions item was the only item failing to load on any factor in their confirmatory factor analysis of the YBOCS. We observed poor test-retest reliability

*To ascertain whether the increase in the size of the correlations at post-test was due to the difference in the sample at that point (not all clients had yet reached the post-test point), we ran an additional set of correlations. We examined the relationship of the YBOCS to the other measures at pretest only for the sample for whom we have post-test data. The correlations were quite comparable to those for the pretest sample as a whole.

for this item. Clinically we find patients have difficulty distinguishing between the degree to which they attempt to resist obsessions and compulsions and their success in doing so, a discrimination the YBOCS asks them to make. We suggest that these items be deleted from the scale unless they are further developed such that their psychometric properties improve.

The two symptom-free interval items (obsessions and compulsions) appeared to add little to the internal consistency or content of the original scale, given the strong relationship between these items and the Time Spent items. We see no benefit to adding these items. Data from the present study do argue for inclusion of the Avoidance investigational item as part of the standard YBOCS score. Avoidance proved reliable and normally distributed and showed good evidence of convergent validity with behavior test and target ratings data. Avoidance has heretofore not been considered as part of the YBOCS despite its clinical importance. It is common for OC clients to avoid threatening situations to reduce their distress and ritualizing (e.g. giving up driving to avoid checking and obsessions about having struck a pedestrian). Further, avoidance has proved to be a strong predictor of outcome in at least one study (Cottraux, Messy, Marks, Mollard & Bouvard, 1993). Although the Avoidance item does not fit neatly within the Obsessions or Compulsions subscales, it could easily be made part of YBOCS Total. We recommend that it be included.

Inter-rater reliability was excellent for both of the subscales and for YBOCS Total, but test-retest reliability coefficients in this study were notably lower than those obtained in previous research by Kim and colleagues (Kim *et al.*, 1990, 1992). Our different procedure may account for the discrepancies. The average test-retest interval was longer in the present study (48.5 vs 7 days), and, unlike Kim *et al.*, we used different raters at the two time points. Thus, our data reflect variability due to inter-rater differences as well as stability of reports and questioning. Given the excellent inter-rater reliability obtained when two raters are provided with the same information via audio-tape, these data suggest that over a long interval, or with independent interviewers who may pursue different lines of questioning, YBOCS data are more variable than had been thought on the basis of previous investigation. Further research on this question would be highly desirable.

With regard to construct validity, the YBOCS and its two subscales correlated moderately strongly with the most commonly used self-report questionnaire measure of OC symptoms, the MOCI. These data conflict with an earlier study by Goodman *et al.* (1989b) showing a low correlation with the MOCI. However, our sample for this analysis was substantially larger and therefore more likely to yield an accurate estimate of the true relationship. Moreover, our findings correspond to those reported by Frost *et al.* (1994) for subclinical Ss.

Previous investigators have not examined the association of the YBOCS with behavioral measures of OCD symptoms. We found consistent evidence of modest relationships between YBOCS Total and Compulsions Scales with behavioral measures (target ratings and behavioral avoidance test). These correlations, although not large, were typically in the range to be expected for heteromethod coefficients with one notable exception: Whereas YBOCS Obsessions demonstrated the predicted correlation to behavior test anxiety, it bore little relationship to target ratings of fear and avoidance. This finding was not expected but may be reasonable: The Fear/Avoidance ratings ask how much clients are distressed by or avoid specific situations that tend to provoke rituals. By contrast, this is only one aspect of the YBOCS Obsessions subscale which also focuses on time spent and interference from obsessions. The modest shared variance between the measures most commonly used in behavioral research (target symptom ratings) and that typically used in pharmacological research (YBOCS) poses a considerable barrier to comparison across treatments. These data argue for the inclusion of both measures in outcome research trials.

The YBOCS showed poor divergent validity in that it was as or more highly correlated with depression than with measures of OCD. Indeed, Goodman *et al.* (1989b) did not find the YBOCS to be consistently more strongly correlated with other OCD measures than with the Hamilton Anxiety and Depression Scales. This lack of divergent validity has also been a consistent problem for other OCD measures, including the MOCI and the Leyton (Emmelkamp, 1988a, b). Our findings indicate this problem does not arise because of the comorbidity of depressive disorders and OCD. When clients with comorbid depression were eliminated from the sample, the problems with divergent validity remained. The relationship between measures of OCD symptoms, anxiety, and depression most likely reflects the distress and debilitation associated with severe OCD. Thus, we would expect a substantial relationship among measures of these symptoms. None the less, from

a psychometric point of view, it is disturbing that the YBOCS in some samples correlates more strongly with moodstate than OCD symptom measures. Thus, like other investigators, we find good evidence for convergent validity but generally poor support for divergent validity *vis à vis* moodstate.

Acknowledgements—Preparation of this manuscript was supported by grant no. MH44190 awarded by the National Institute of Mental Health to Gail Steketee and Dianne Chambless. The authors thank interviewers Martha Gillis, M.A., Whitney Wykoff, M.Ed., Hope Worden, MSW, and Cheryl Sheffler Rubenstein, M.A. whose assistance was instrumental in completing this project. We are grateful to Dr Wayne Goodman who provided training in YBOCS interviewing and kindly responded to questions regarding ratings on this measure.

REFERENCES

- Cottraux, J., Messy, P., Marks, I. M., Mollard, E. & Bouvard, M. (1993). Predictive factors in the treatment of obsessive-compulsive disorders with fluvoxamine and/or behaviour therapy. *Behavioural Psychotherapy*, 21, 45–50.
- Danyko, S. J., McKay, D. R. & Neziroglu, F. A. (1993, November). Factor structure and construct integrity of the Yale–Brown Obsessive Compulsive Scale. Paper presented at the meeting of the Association for Advancement of Behavior Therapy, Atlanta, Ga.
- Derogatis, L. R. (1977). *SCL-90-R: Administration, scoring and procedures manual I*. Baltimore, Md: Clinical Psychometrics Research.
- Emmelkamp, P. M. G. (1988a). Leyton Obsessional Inventory. In Hersen, M. & Bellack, A. S. (Eds), *Dictionary of behavioral assessment techniques* (pp. 287–291). New York: Pergamon Press.
- Emmelkamp, P. M. G. (1988b). Maudsley Obsessional–Compulsive Inventory. In Hersen, M. & Bellack, A. S. (Eds), *Dictionary of behavioral assessment techniques* (pp. 294–296). New York: Pergamon Press.
- Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Freund, B. (1986). Comparison of measures of obsessive-compulsive symptomatology. Unpublished doctoral dissertation, Southern Illinois University at Carbondale. Ann Arbor, Mich: University Microfilms International.
- Frost, R. O., Steketee, G., Krause, M. S. & Trepanier, K. L. (1994). The relationship of the Yale–Brown Obsessive Compulsive Scale (YBOCS) to other measures of obsessive compulsive symptoms in a nonclinical population. Manuscript under editorial review.
- Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., Heninger, G. R. & Charney, D. S. (1989a). The Yale–Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Archives of General Psychiatry*, 46, 1006–1011.
- Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Delgado, P., Heninger, G. R. & Charney, D. S. (1989b). The Yale–Brown Obsessive Compulsive Scale. II. Validity. *Archives of General Psychiatry*, 46, 1012–1016.
- Hodgson, R. J. & Rachman, S. J. (1977). Obsessional-compulsive complaints. *Behaviour Research and Therapy*, 15, 389–395.
- Kim, S. W., Dysken, M. W. & Kuskowski, M. (1990). The Yale–Brown Obsessive–Compulsive Scale: A reliability and validity study. *Psychiatry Research*, 34, 99–106.
- Kim, S. W., Dysken, M. W. & Kuskowski, M. (1992). The Symptom Checklist-90: Obsessive–Compulsive Subscale: A reliability and validity study. *Psychiatry Research*, 41, 37–44.
- Spitzer, R., Williams, J., Gibbon, M. & First, M. (1990a). *Structured Clinical Interview for DSM-III-R—Patient Edition (With Psychotic Screen) Version 1.0* Washington, DC.: American Psychiatric Press.
- Steketee, G. (1993). *Treatment of obsessive-compulsive disorder*. New York: Guilford.