2019-05-01

Y-BOCS Factor Structure Analysis and Calculation of Measurement and Structural Invariance Between Genders

Sean B. Vanhille
Brigham Young University

Y-BOCS Factor Structure Analysis and Calculation of Measurement and Structural Invariance Between Genders

Sean B. Vanhille

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Scott A. Baldwin, Chair
Michael J. Larson
Scott R. Braithwaite

Department of Psychology

Brigham Young University

ABSTRACT

Y-BOCS Factor Structure Analysis and Calculation of Measurement and Structural Invariance Between Genders

Sean B. Vanhille
Department of Psychology, BYU
Master of Science

The Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) is considered the "gold standard" for measuring symptoms for Obsessive-Compulsive Disorder (OCD) due to the high reliability and validity of the measure. Originally, the Y-BOCS was divided into Obsessive and Compulsive factors; however, literature on the factor structure of the Y-BOCS is inconsistent. Models range from one global factor to different interpretations of bi-factor models to three-factor models. Inconsistencies between models may be attributed to sampling error, including participants with subclinical OCD in some samples, and measurement error. In addition, many researchers treat the Y-BOCS measurement as an interval or ratio scale when it likely reflects ordinal measurement.

Our paper has two primary aims. First, we compare the fit of the models proposed in the literature using a large sample from multiple sites of patients diagnosed with OCD. We also evaluate how the models can be improved and whether those improvements show evidence for convergent validity. We treat the Y-BOCS observations as ordinal data. Second, we evaluate measurement and structural invariance between genders. Additionally, we examine convergent validity of the factor structure of the best fitting model with subscales of the OCI-R.

Data from five separate samples were combined into one dataset with 288 total participants all formally diagnosed with OCD. We selected several Y-BOCS factor models from the literature and used confirmatory factor analysis to evaluate goodness of fit indices on our pooled sample. Only one model approached acceptable goodness of fit indices. We considered the factors in this model and proposed a new factor model with a global factor (OCD) and two sub-factors (Obsessions and Resistance to Symptoms). Our model exhibited the highest goodness of fit indices which we further improved with modifications to our factor model. On invariance analyses, our model exhibited measurement invariance between genders and partial structural invariance. Additionally, the latent factors of our model exhibited convergent validity with all of the OCI-R subscales (except Ordering).

Our model exhibited stronger goodness of fit indices with our data than existing models in the Y-BOCS literature and measurement invariance and partial structural invariance between genders. We recommend that future studies replicate the efficacy of our factor model using the Y-BOCS as an ordinal measurement.

Keywords: invariance, factor structure, Obsessive-Compulsive Disorder

ACKNOWLEDGEMENTS

I would like to thank Dr. Baldwin for the weekly meetings in which he patiently answered my many questions. His kind encouragement helped me to problem solve diligently and to write consistently. I also appreciate Dr. Larson and Dr. Braithwaite for their support and feedback on this project. Lastly, I would like to express gratitude for my family and their endless love and support, especially when I am stressed and exhausted.

TABLE OF CONTENTS

List of Tables

List of Figures

Y-BOCS Factor Structure Analysis and Calculation of Measurement and Structural Invariance Between Genders

Obsessive-Compulsive Disorder (OCD) is an anxiety disorder that involves intrusive, distressing thoughts (obsessions) and compulsive behaviors that reduce the distress (compulsions). The Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) is a 10-item clinician-rated scale that assesses the presence and severity of obsessive and compulsive symptoms (Goodman et al., 1989a, b). The Y-BOCS is used extensively in clinical and research settings and is considered the "gold standard" in assessing OCD symptom severity (Antony, Orsillo, & Roemer, 2001). The content of the 10 items cluster around the constructs of obsessions (items 1-5) and compulsions (items 6-10). Responses are made on a 5-point Likert Scale. Items on the Y-BOCS cover various domains: time spent (in hours) ranging from *none* to *nearly constant occurrence*, interference (in social or occupational performance) ranging from *none* to *incapacitating*, distress (degree of disturbance) ranging from *none* to *near constant and disabling stress*, resistance (effort to resist or disregard) ranging from *try to resist all the time* to *completely and willingly yield to all obsessions*, and degree of control ranging from *complete control* to *obsessions are completely involuntary/rarely able to even momentarily delay action*. The sum of the 10 items produces an overall symptom score whereas the sum of the 5 obsession (or compulsion) items produces a subscale score. Higher scores indicate higher levels of impairment.

Psychometric Evidence

Goodman and colleagues (1989a) developed the Y-BOCS using a clinical sample and the instrument demonstrated high levels of reliability both initially and in several subsequent studies. For example, inter-rater reliability between 4 raters was 0.98 for the total score, 0.97 for the obsessions subscale, and 0.96 for the compulsions subscale (Goodman et al., 1989a). Inter-rater reliability between the 4 raters for individual items ranged from 0.86 to 0.97. In a meta-analysis studying the reliability of the Y-BOCS, the average inter-rater reliability was 0.92 (López-Pina et al., 2015). Internal consistency for 4 raters on the Y-BOCS averaged to 0.89 (Goodman et al., 1989a). In a meta-analysis analyzing the internal consistency of the Y-BOCS, the mean coefficient alpha was 0.87 (López-Pina et al., 2015). Correlations between the total score on the Y-BOCS and each item ranged from 0.36 to 0.77 (Goodman et al., 1989a). In a meta-analysis analyzing the test-retest reliability of the Y-BOCS, the average estimate was 0.85 (López-Pina et al., 2015).

Goodman and colleagues also showed that the Y-BOCS exhibited high levels of validity and sensitivity to symptom severity (1989b). Total scores on the Y-BOCS exhibited strong correlations with other measures of OCD such as the National Institutes of Mental Health Global Obsessive-Compulsive Scale (NIMH-OC), $r = .67$ $p < .001$, and a modified form of the Clinical Global Impression Scale for Global Severity of Obsessive-Compulsive Scale (CGI-OCS), $r = .74$ $p < .001$ (Goodman et al., 1989b). Additionally, total YBOCS scores were moderately correlated with the Maudsley Obsessive-Compulsive Inventory (MOCI) Total score both pre- and post-treatment at .43 and .55, $p < .005$, respectively (Woody, Steketee, & Chambless, 1995). Patients with OCD scored higher on the Y-BOCS than patients with anxiety disorders (Rosenfeld, Dar, Anderson, Kobak, & Greist, 1992). However, subsequent research indicated poor discrimination between OCD and depressive and anxiety symptoms (Goodman et al., 1989b; Taylor, 1995; Woody et al., 1995). Total Y-BOCS scores showed sensitivity to symptom severity as patients treated with medication reported reduced OCD symptoms. Total scores significantly decreased in groups treated with medication by 42% when compared to those in a placebo group (Goodman et al., 1989b).

Factorial Validity

Despite being considered the "gold standard" in the field for measuring OCD, research involving the factorial validity of the Y-BOCS is inconsistent (Anholt et al., 2010). Fals-Stewart (1992) showed that the Y-BOCS could load on a single, global factor which indicated OCD impairment. However, the inclusion of six investigational items in addition to the original 10 items weaken the generalizability of these results. Others have shown that 2- or 3-factor models exhibit clustering on factors such as disturbance (items 2, 3, 7, 8) and symptom severity (items 1, 4, 5, 6, 9, 10; Amir, Foa, & Coles, 1997); resistance/control (items 4, 5, 9, 10) and symptom severity (items 1, 2, 3, 6, 7, 8; Deacon & Abramowitz, 2005); and obsession severity (items 1, 2, 3, 5), compulsion severity (items 6, 7, 8, 10), and resistance to symptoms (items 4, 9; Kim, Dysken, Pheley, & Hoover, 1994).

Varying sample composition and statistical techniques may have produced the differing factor structures reported. Table 1 summarizes demographic and statistical techniques exploring the factor structure of the Y-BOCS in the current literature. The samples used in each study included outpatients (Anholt et al., 2010; Fals-Stewart, 1992; McKay, Danyko, Neizroglu, & Yaryura-Tobias, 1995), inpatients (Arrindell, de Vlaming, Eisenhardt, van Berkum, & Kwee, 2002; Moritz et al., 2002), individuals within clinical trials (Kim et al., 1994; Storch et al., 2005), and undergraduates who did not have an OCD diagnosis (Garnaat, & Norton, 2010). Because the Y-BOCS was originally shown to be sensitive to a change in OCD symptoms, these previously mentioned groups may experience symptoms differently (Goodman et al., 1989b).

Table 1.

Past Research Sample and Measurement

	N	Female	Age	Composition	Dx with OCD	Measurement Type
Fals-Stewart*	193	58%	30.5 (7.9)	Outpatient	DSM-III	Interval (PCA-promax rotation)
Goodman	300	Not reported	Not reported	“Patients”	DSM-III	Interval
Kim	214	61%**	35.4 (10.3)	Clinical Trials	DSM-III	Interval (PCA-varimax rotation)
Amir	Two 202 samples	51%	36.4 (12.3)	Inpatient & Outpatient	DSM-III	Interval (CFA)
Deacon	100	49%	35.8 (12.2)	Outpatient	DSM-IV	Interval (CFA then PCA-oblique rotation)
McKay	83	47%	43.0 (8.8)	Outpatient	DSM-III	Interval (CFA)
Anholt	544 (split in half)	62.7%	37.1 (11.06)	Outpatient	DSM-IV	Both halves: PCA-promax rotation; CFA
Arrindell	65	62.9%	34.0 (9.0)	Inpatient	DSM-III	Multiple Group Method (MGM) confirmatory analysis
Moritz	109	53.2%	33.2 (9.9)	Inpatient		PCA-varimax rotation
Storch	131	53.4%	34.2 (11.3)	Clinical Medication Trial; Outpatient Clinic	DSM-III or DSM -IV	CFA
Our Proposed Approach	288	53%	31.41 (11.88)	Outpatient	DSM-IV	Ordinal (CFA)

Note: *Fals-Stewart (1992) used a 16-item YBOCS rather than the traditional 10-item version. **In Kim’s (1994) study, the original sample was 238 of which 23 withdrew prematurely and gender percentage was not reported on the final sample.

Anholt and colleagues (2010) noted the lack of consistency in factor analytic methods used with the Y-BOCS. Some studies used exploratory factor analysis and varied with respect to the rotation methods (Fals-Stewart, 1992; Kim et al., 1994; Moritz et al., 2002) and employing different rotations in these exploratory factor analyses. For example, some researchers employed a varimax rotation (Kim et al., 1994; Moritz et al., 2002). A varimax rotation is a type of orthogonal rotation with the underlying assumption that factors are uncorrelated; however, in the social sciences, few factors are uncorrelated (Costello & Osborne, 2005). Specifically, in regard to the Y-BOCS, latent factors represent aspects of the diagnosis of OCD and, subsequently, likely are correlated to at least a small degree. Consequently, the use of an oblique rotation (such as a promax rotation) would likely provide a more accurate representation of the correlation between factors. Several researchers performed this type of rotation in their analysis (Anholt et al., 2010; Deacon & Abramowitz, 2005; Fals-Stewart, 1992)

Other research in the field utilized confirmatory factor analysis (Amir et al., 1997; Arrindel et al., 2002; Deacon and Abramowitz, 2005; McKay et al., 1995; McKay, Neziroglu, Stevens, & Yaryura-Tobias, 1998; Storch et al., 2005). In one of the analyses, the sample size was relatively small with 83 participants. In another analysis, the CFA exhibited poor fit so a principal component analysis (PCA) was then performed on the data with an oblique rotation. However, PCA includes both shared and unique variance while factor analysis focuses on shared variance to help reveal how latent variables cause covariation between observed variables (Costello & Osborne, 2005).

Ordinal Measure

Much of the research to date treats the Y-BOCS as an interval level of measure (see Garnaat & Norton, 2010, for an exception). The 5-point Likert scale of the Y-BOCS may better reflect an ordinal level of measurement. Scales on an ordinal level of measurement exhibit a ranking without equivalent intervals between scores while an interval scale exhibits both ranking and equivalent intervals between scores (Stevens, 1946). The theoretical "distance" between *mild* to *moderate* may not be the same as *severe* to *extreme* (Boone & Boone, 2012). The anchors for each item on the Y-BOCS are more consistent with an ordinal scale than a Likert scale. The pairing of questions between obsessions and compulsions on various constructs (e.g., distress or resistance) may further impair the consistency of the scaling of responses on the Likert scale (Boone & Boone, 2012). This pairing would also fit with Stevens' (1946) description of "relative rank-ordering" which, he stated, was the level of measurement of many psychological tools. This potentially arbitrary decision of level of measurement is not solely in the field of psychology alone; a review of medical research indicated that up to one-fifth of articles published utilized ordinal data and analyzed without addressing that level of measurement (Forrest & Andersen, 1986).

This question of level of measurement becomes relevant with performing factor analysis calculations. Maximum likelihood (ML) is the most common method of estimation in CFA and often performed with EFA (Baglin, 2014; Flora & Curran, 2004). This method has the assumption that the observed variable is continuous and normally distributed. However, this assumption is not met when observed data is discrete which results in challenges to fit CFA models with ordinal data, particularly when the number of observed categories five or fewer (Flora & Curran, 2004). Consequently, using ML to estimate factor models with Y-BOCS data is likely flawed. Alternatively, using a method of estimation such as weighted least squares (WLS) results in a more accurate fit due to allowing for dichotomous, ordered categorical, or continuous observed variables (Flora & Curran, 2004). More specifically, weighted least square mean and variance adjusted estimation (WLSMV) is designed for categorical observed data such as dichotomous or ordinal data (Li, 2016). The major underlying assumption for this estimator is that while the observed data may not be continuous, the latent variable exists on a normal distribution. WLSMV tends to outperform other estimators when using ordinal data as factor loadings are typically unbiased, more accurate, and more precise (Li, 2016).

Measurement Invariance

The DSM-5 reports that OCD onset often varies between gender with an earlier onset in males than in females (American Psychiatric Association, 2013). This onset difference is reflected in the literature (Castle, Deale, & Marks, 1995; Mathis et al., 2011). Clinical features of OCD show variation across gender as well (Labad et al., 2008; Mathis et al., 2011). Males typically exhibit greater social impairment (e.g., 2/3 remain single compared to only 1/3 of females), more sexual-religious and aggressive symptoms (F:M adjusted *OR* = .041), and greater comorbidity with tic and substance use disorders. Females typically present more contamination/cleaning symptoms (F:M adjusted *OR* = 2.05) and greater comorbidity rates with eating and impulse-control disorders such as skin-picking. Since gender differences exist between genders with OCD, then using a tool with a factor structure that measures constructs similarly between genders would be valuable. Fortunately, we can use measurement invariance techniques to evaluate whether the Y-BOCS is psychometrically equivalent for men and women.

Measurement invariance is defined as whether a latent variable, or measured construct, is equivalent under different conditions (Horn & McArdle, 1992). Measurement invariance analyses indicate the level of similarity across groups of the proposed latent factor(s) measured (Baldwin, 2019; Reise, Widaman, & Pugh, 1993; Steenkamp & Baumgartner, 1998). Analyses to support measurement invariance across groups are done in stepwise fashion by comparing configural, metric, scalar, and residual invariance models. This calculation systematically constrains aspects of the factor structure between groups and assesses any potential change in goodness of fit indices from one calculation to the next (i.e., configural to metric, metric to scalar, scalar to residual). A commonly used index for change in fit is the Likelihood Ratio Test (LRT), which calculates the difference in χ^2 values between the two models (Cheung & Rensvold, 2002). The null hypothesis is that two models have identical fit and a significant difference in χ^2 values indicates that invariance was not met. For example, if the LRT indicates a significant difference between the metric invariance model and the scalar invariance model, then the measure is not invariant at the scalar level even though it may be invariant at the metric level.

Configural invariance is the first step in the process of calculating measurement invariance; it indicates that the pattern of factor loadings between the two models (e.g., gender, race, etc) is equivalent in both groups (Steenkamp & Baumgartner, 1998). Establishing configural invariance is crucial for meaningful comparisons in later steps (Lance & Vandenberg, 2002). Metric invariance is calculated by comparing the configural model to a metric model (Baldwin, 2019). Metric invariance models include constraints to the factor loadings from latent factors to item means between two groups (Steenkamp & Baumgartner, 1998). If support for metric invariance occurs, then this means that the item means on each item is due to similar factor loadings between a latent variable and observed item means (Steenkamp & Baumgartner, 1998). Scalar invariance models include additional constraints of item intercepts from the metric model (Steenkamp & Baumgartner, 1998). If support for scalar invariance occurs, then the item intercepts are statistically similar or that the two groups respond similarly to each item; a difference in item means is due to the latent factor measured rather than measurement bias (Baldwin, 2019). Residual invariance models include additional constraints to item residuals

(Baldwin, 2019). If support for residual invariance occurs, then this means that differences between groups with observed means and variances is due to differences in the latent factor means and variances (Widaman & Reise, 1997).

To the best of our knowledge, only one published study has calculated measurement and structural invariance using the Y-BOCS and this study showed that the Y-BOCS is invariant among races (Garnaat & Norton, 2010). Of note, they treated the Y-BOCS as an ordinal measure rather than as continuous data. Garnaat and Norton (2010) showed that the original two-factor model proposed by Goodman and colleagues (1989a, b) exhibited measurement invariance when comparing White group to each Asian and Hispanic group. However, the White group compared to the Black group did not result in invariance with underestimations of interference, distress, and resistance due to obsessions in the Black group.

Aims

Our paper has two primary aims. First, we compare the fit of the models proposed in the literature using a large sample from multiple sites of patients diagnosed with OCD. We also evaluate how the models can be improved and whether those improvements show evidence for convergent validity. We treat the Y-BOCS observations as ordinal data. Second, we evaluate measurement and structural invariance between genders. Additionally, we examine convergent validity of the factor structure of the best fitting model with subscales of the OCI-R.

Method

Procedure and Participants

Data for the current analyses came from five separate previously-conducted studies. Four of the studies were published (Storch, Abramowitz, & Keeley, 2009; Storch et al., 2006, 2007, 2008) and one was unpublished data on cognitive control deficits and dysfunction associated with OCD (Larson, 2012). The measurements from these studies occurred during baseline sessions before any pharmacological or psychotherapeutic interventions began. All of the participants were given a primary diagnosis of OCD based on an initial a clinical interview by a licensed clinical psychologist, licensed psychiatrist, physician, or graduate student supervised by a licensed professional. These diagnoses were subsequently confirmed by either the Anxiety Disorder Interview Schedule for DSM-IV (ADIS: Brown, Barlow, & DiNardo, 1994) or the Structured Clinical Interview for the DSM-IV (SCID-IV: First & Gibbon, 2004). The same psychologist, psychiatrist, or graduate student who diagnosed the participant also administered the Y-BOCS to that participant according to standardized administration procedures (Goodman et al., 1989a, b). All but one of the samples also included additional measures during the baseline sessions (see details below).

After pooling across the five sample, the final sample included 288 participants (140 female) ranging in age from 16 to 79 (*M* = 31.41, *SD* = 11.88). Not all demographic information can be provided for the pooled sample due to variability data collection. For example, not all sites measured the time of onset of OCD or level of education. Available information on each sample is presented below.

Sample 1 (Storch et al., 2009) included 95 participants (46 female) with an age range from 16 to 62 years (*M* = 34.63, *SD* = 11.44) and education ranging from 8 to 24 years (*M* = 15.25, *SD* = 3.09). 95.8% participants identified as White with 2.1% identified as Black and 2.1% identified as "Other." No other measures were administered to determine additional diagnoses, but 33.7% of participants self-reported other diagnoses in addition to their primary diagnosis of OCD.

Sample 2 (Larson, 2012) included 25 participants (12 female) with an age range from 18 to 53 years (*M* = 24.52, *SD* = 7.03) and education ranging from 12 to 18 years (*M* = 14.82, *SD* = 1.64). All of the participants identified as White. Participants completed both the Y-BOCS and the Obsessive-Compulsive Inventory Revised (OCI-R). They also completed measures to determine diagnoses in addition to OCD that covered domains for emotional concerns, learning, memory, and executive functioning. Other diagnoses included two with panic disorder, one with social phobia, five with generalized anxiety disorder (GAD), and five with major depression.

Sample 3 (Storch et al., 2008) included 62 participants (29 female) ranging from age 18 to 61 (*M* = 30.27, *SD* = 12.16). 98.4% of participants identified as White with 1.6% of participants identified as Black. No educational information was collected. Participants completed both the Y-BOCS and the OCI-R. They also completed measures to determine diagnoses in addition to OCD such as screening tools and emotional concerns. Other diagnoses included 12 with panic disorder, five with agoraphobia, 17 with social phobia, 22 with GAD, 24 with major depression, and 6 with dysthymia.

Sample 4 (Storch et al., 2007) included 29 participants (15 female) ranging from age 18 to 53 (*M* = 28.79, *SD* = 9.23). 86.2% of participants identified as White with 3.4% of participants identified as Black, 3.4% of participants identified as Asian, and 3.4% of participants identified as “Other.” No educational information was collected. Participants completed both the Y-BOCS and the OCI-R. They also completed measures to determine diagnoses in addition to OCD such as screening tools and emotional concerns. Other diagnoses included three with panic disorder, one with agoraphobia, one with GAD, six with major depression, and five with dysthymia.

Sample 5 (Storch et al., 2006) included 77 participants (38 female) ranging from age 18 to 65 (*M* = 31.55, *SD* = 13.18). 85.7% of participants identified as White with 3.9% of participants identified as Black, 6.5% of participants identified as Hispanic, and 3.9% of participants identified as “Other.” No educational information was collected. Participants completed the Y-BOCS and other measures addressing depression and tics. Other diagnoses included three with panic disorder, six with social phobia, sixteen with GAD, thirty-three with major depression, and four with dysthymia.

Measures

Y-BOCS. The Y-BOCS (Goodman et al., 1989a, b) is administered by a trained clinician or graduate student to assess obsessive-compulsive symptoms on a 5-point Likert scale on each of the ten questions. Obsessions and compulsions are rated with one question each identifying distress, frequency, interference, resistance, and control of symptoms. Scores can be totaled for obsessions, compulsions, and a total score which combines the obsession and compulsion scores. The Y-BOCS exhibits both high validity and reliability as discussed above.

OCI-R. The Obsessive-Compulsive Inventory--Revised (OCI-R) self-report measure contains 18 items with 6 subscales: checking, hoarding, neutralizing, obsessing, ordering, and washing (Foa et al., 2002). Each question is rated on a 5-point scale to assess the level of distress experienced by individuals within the past month which ranges from “Not at all” to “Extremely”. The OCI-R provided three benefits beyond its parent scale the Obsessive-Compulsive Inventory (OCI): reduced redundancy, minimized overlap between subscales, and improved scoring ease (Foa et al., 2002). Items can be totaled by subscale and overall score.

The OCI-R also exhibits good psychometric properties (Abramowitz & Deacon, 2006; Foa et al., 2002; Hajcak, Huppert, Simons, & Foa, 2004; Huppert et al., 2007). Foa and colleagues (2002) developed the shortened, revised instrument using clinical samples which included those diagnosed with OCD, other anxiety disorders, and normal controls. They reported that internal consistency of the measure with those diagnosed with OCD ranged from .81 to .90 (Foa et al., 2002). Other studies with a clinical sample showed similarly good internal consistency values of .83 and .84 (Abramowitz & Deacon, 2006; Huppert et al., 2007). In a college sample, the OCI-R exhibited excellent internal consistency of .88 (Hajcak et al., 2004). Correlations among subscales ranged from .31 to .57 and correlations from between the subscales and total score ranged from .63 to .80 (Foa et al., 2002). Test-retest reliability among OCD patients ranged from .74 to .91 (Foa et al., 2002). Another study with a clinical sample showed a similarly good test-retest reliability of .70 (Huppert et al., 2007). Total score correlations between the OCI-R and the OCI were .98 with all individual subscale correlations above .90 except for the Neutralizing subscale which was .74 (Foa et al., 2002).

The OCI-R also exhibits good validity. The convergent validity between the total OCI-R score and total Y-BOCS score indicated a moderate correlation of .53 (Foa et al., 2002). Another study of patients with OCD showed a correlation of .41 between the total OCI-R and Y-BOCS scores (Abramowitz & Deacon, 2006). The correlation between the total OCI-R score and the NIMH-OC and the MOCI were .66 and .85, respectively (Foa et al., 2002). In a college sample, the total OCI-R and MOCI total scores exhibited a moderate correlation of .56 (Hajcak et al., 2004). Divergent validity between the OCI-R and each the Beck Depression Inventory (BDI) and Hamilton Rating Scale for Depression (HRSD) were moderate correlations of .70 and .58, respectively (Foa et al., 2002). Additionally, ROC analyses indicated that the OCI-R exhibits good sensitivity and specificity between patients with OCD and anxiety and non-anxiety controls

(Foa et al., 2002). Another study with clinical samples showed good discrimination between patients with OCD and GAD using subscales of the OCI-R (Huppert et al., 2007).

Data Analyses

To address the first aim, we assessed the fit of previously proposed Y-BOCS factor models using Mplus 8. For reasons discussed above, we treated all Y-BOCS responses as ordinal data. We estimated five models based on the previous literature: (a) 1 global factor OCD (Fals-Stewart, 1992); (b) 2 factors Obsessions (items 1-5) and Compulsion (items 6-10) (Goodman et al., 1989a, b); (c) 2 factors Disturbance (items 2, 3, 7, 8) and Symptom Severity (items 1, 4, 5, 6, 9, 10) (Amir et al., 1997); (d) 2 factors Symptom Severity (items 1, 2, 3, 6, 7, 8) and Resistance/Control to Symptoms (4, 5, 9, 10) (Deacon & Abramowitz, 2005); and (e) 3 factors Resistance to Symptoms (items 4, 9), Severity of Obsessions (items 1, 2, 3, 5), and Severity of Compulsions (6, 7, 8, 10) (Kim et al., 1994). We utilized χ^2 goodness-of fit, RMSEA, CFI, and WRMR indices to determine goodness of fit. For the χ^2 goodness-of fit-index, the null hypothesis states that the covariance matrix of the sample and the model are the same (Cheung & Rensvold, 2002). Consequently, a failure to reject the null hypothesis indicates that the two matrices are statistically similar, and the model is considered a good fit. However, despite its wide use, large sample sizes complicate interpretation, so this statistic is often used in conjunction with other indices (Cheung & Rensvold, 2002). For RMSEA, values at 0.05 or less represent “close fit,” values at 0.08 or less represent “reasonable fit,” and values greater than 0.10 represent “unacceptable fit” (Browne & Cudeck, 1993). For CFI, values at and above 0.95 indicate good fit (Hu & Bentler, 1999). For WRMR, which was developed for use with ordinal data, values below 1.00 represent good fit (DiStefano, Liu, Jiang, & Shi, 2018). Because none of the models exhibited excellent fit with our pooled sample, we proposed a new model based on the current models in the literature.

To address our second aim, we fit models to evaluate aspects of measurement invariance with our proposed model. Measurement invariance analyses indicate the level of similarity across groups of the proposed latent factor(s) measured (Baldwin, 2019; Reise et al., 1993; Steenkamp & Baumgartner, 1998). We analyzed measurement invariance across gender by calculating in stepwise fashion configural, metric, scalar, and residual invariance. These analyses systematically add constraints to aspects of the factor structure between genders and assesses any potential change in goodness of fit indices from one calculation to the next by using the Likelihood Ratio Test (LRT) which calculates the difference in χ^2 values between the two models compared (Cheung & Rensvold, 2002). A significant difference in χ^2 values indicates that invariance was not met. For example, if the LRT indicates a significant difference between the metric invariance model and the scalar invariance model, then the measure is not invariant at the scalar level. This result would indicate that the item intercepts are not statistically similar or that the two groups do not respond similarly to each item; a difference in item means is not due only to the differences in latent factor means and variances between groups but may be due to measurement bias (Baldwin, 2019). Table 2 displays what aspects of the model are allowed to be freely estimated and which are constrained when calculating measurement invariance.

Table 2.

Constraints at Each Step of Calculating Measurement Invariance

	Configural		Metric		Scalar		Residual	
	Men	Women	Men	Women	Men	Women	Men	Women
Factor Structure	F	F	F	F	F	F	F	F
Factor Loading	F	F	F	C	F	C	F	C
Item Intercepts	F	F	F	F	F	C	F	C
Residual Error	I	I	I	I	I	I	I	F
Factor Variance	I	I	I	F	I	F	I	F
Factor Means	I	I	I	I	I	F	I	F

Note: F = Free (freely estimated values in the model). C = Constrained (constrained estimates in the model). I = Identification (constrained for identification of the model).

After calculating measurement invariance, we then evaluated structural invariance. This analysis indicates whether the factor means and/or variances are statistically similar between groups (Baldwin, 2019). For example, if a model meets structural invariance, then the latent factor mean and variance of the construct "obsessions" are statistically similar between men and women. If not, then men may experience more obsessions than women or experience obsessions in a wider range. Table 3 displays what aspects of the model are allowed to be freely estimated and which are constrained when calculating structural invariance.

Table 3.

Constraints at Each Step of Calculating Structural Invariance

	Factor Variances		Factor Means	
	Men	Women	Men	Women
Factor Structure	F	F	F	F
Factor Loading	F	C	F	C
Item Intercepts	F	C	F	C
Residual Error	I	I	I	I
Factor Variance	I	C	I	I
Factor Means	I	F	I	I

Note: F = Free (freely estimated values in the model). C = Constrained (constrained estimates in the model). I = Identification (constrained for identification of the model).

The final analysis determined convergent validity of the factor structure of the best fitting model with the OCI-R. As noted previously, the total OCI-R score exhibits good convergent validity with the total Y-BOCS score (Foa et al., 2002). Additionally, in a clinical sample, the OCI-R subscales exhibited mild-to-moderate correlations with the total score for Obsessions, Compulsions, and overall total score of the Y-BOCS as well as good discrimination between patients with OCD and patients with other anxiety disorders (Abramowitz & Deacon, 2006). Due to the good psychometric properties of the OCI-R total score and subscale scores, we analyzed if the latent factors of the best fitting model significantly predicted OCI-R total and subscale scores. If the measured means of the latent factors of the best fitting model significantly predict subscale or total scores on the OCI-R, then this result would provide additional support and validity of the factor structure of the Y-BOCS.

Results

Descriptive Statistics

Figure 1 displays how often clinicians used the response options for each Y-BOCS question for the N = 288 (48.6% female) participants. The spread between ratings appeared good with the exception of relatively few participants endorsing the lowest rating of severity. We expected this exception because participants diagnosed with OCD were recruited for each study which would result in a bias towards greater symptom severity.

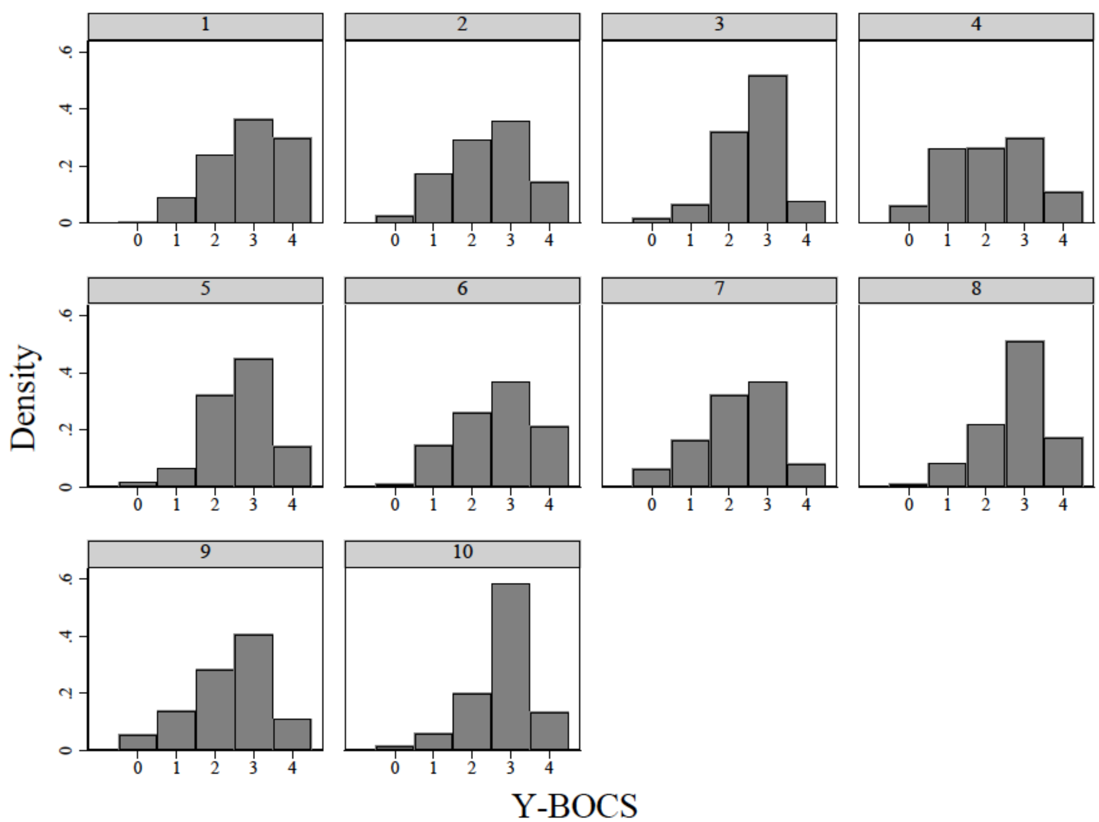

Figure 1. Participant response distribution for each Y-BOCS item.

Model Comparison

We calculated goodness of fit indices for each of the five previously mentioned models selected from the literature. As noted previously, we used conventional standards to determine fit with a nonsignificant χ^2 test indicating good fit, RMSEA values less than 0.05 representing good fit, CFI values at and above 0.95 indicating good fit, and WRMR values below 1.0 representing good fit (Browne & Cudeck, 1993; Cheung & Rensvold, 2002; DiStefano et al., 2018; Hu & Bentler, 1999). The fit indices for each of the models are shown in Table 4.

Table 4.

Fit Indices for Existing Y-BOCS Models in the Literature, Our Proposed Model, and Our Modified Model

	Goodman	Fals-Stewart	Kim	Amir	Deacon	Our Model	Our Modified Model
χ^2 (df)	347.0*** (34)	344.2*** (35)	299.0*** (32)	297.6*** (34)	148.3*** (34)	117.6*** (29)	69.5*** (25)
RMSEA	.18	.18	.17	.16	.11	.11	.08
CFI	.85	.85	.87	.87	.95	.96	.98
WRMR	1.84	1.88	1.66	1.75	1.20	.91	.69

Note: Our model showed the best fit with CFI and WRMR values indicating good fit. Our modified model showing further improvement. *** $p < .001$

Due to the poor fit of the current selected models in the literature, we proposed a new, bifactor model based on these models. A bifactor model contains a general factor that accounts for the shared variance among all facets or components of the factor while the specific factors account for certain unique aspects above and beyond the general factor (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012; Gibbons & Hedeker, 1992). This approach requires an examination of total scores and subscale scores to determine if subscales or specific factors provide additional, useful information or account for variance beyond the total score or general factor. Using a bifactor approach may lead to the inclusion or exclusion of a subscale or specific factor depending on the strength of the relationship with desired dependent variable or even reveal specific factors with an inverse relationship with one another and a dependent variable (Chen et al., 2012). Our proposed model contained a global or general factor of OCD which loaded onto each of the items (like the Fals-Stewart model, 1992). We used this global factor as a way to capture the unitary aspect of OCD and account for the shared variance of all aspects of OCD. The model also contained two other specific to capture aspects over and above the general factor OCD: Obsessions (items 1-3) and Resistance (items 4-5, 9-10). Obsessions is a core aspect of OCD to make a diagnosis, so we formed a latent factor to reflect these symptoms. The Obsessions factor in our model is similar to the Obsessions factor in the Goodman model (1989a) except for items 4 and 5 which loads onto the Resistance factor instead. The Resistance factor in our model loads the same as the Resistance factor in the Deacon model. The Deacon model showed the best fit when compared to the other models in the literature we selected so we included their specific factor of Resistance in our model.

Our proposed model exhibited the best fit compared to the other selected models from the current literature (RMSEA = .11, CFI = .96, WRMR = .91). We proceeded to make some adjustments to our model to further improve fit based on conceptual considerations. We allowed covariation of residual errors between items 1 and 6 as both of those items address time (with obsessions and compulsions, respectively). We allowed covariation of residual errors between items 2 and 7 as both of those items address interference (with obsessions and compulsions, respectively). We allowed covariation of residual errors between items 3 and 8 as both of those items address distress (with obsessions and compulsions, respectively). The pairing of items 4 and 9 or items 5 and 10 were addressed with the latent factor Resistance rather than through covariation of errors. These 3 adjustments with covariation of errors further improved fit for our model (RMSEA = .08, CFI = .98, WRMR = .69). Factor loadings with 95% CI and standardized covariances for between items for our modified proposed model are displayed in Tables 5.

Table 5.

Our Model Standardized Factor Loadings

	General Factor		Specific Factor		Specific Factor	
	Factor Loadings	95% CI	Factor Loadings	95% CI	Factor Loadings	95% CI
Items	OCD		Obsessions		Resistance	
YBOCS 1	0.51 (0.06)	[0.40, 0.62]	0.63 (0.08)	[0.49, 0.78]		
YBOCS 2	0.58 (0.05)	[0.49, 0.68]	0.49 (0.07)	[0.35, 0.63]		
YBOCS 3	0.64 (0.05)	[0.55, 0.73]	0.49 (0.07)	[0.36, 0.62]		
YBOCS 4	0.39 (0.06)	[0.27, 0.51]			0.43 (0.07)	[0.29, 0.56]
YBOCS 5	0.50 (0.05)	[0.40, 0.60]			0.35 (0.06)	[0.24, 0.46]
YBOCS 6	0.74 (0.04)	[0.67, 0.81]				
YBOCS 7	0.88 (0.03)	[0.84, 0.94]				
YBOCS 8	0.76 (0.03)	[0.71, 0.83]				
YBOCS 9	0.39 (0.06)	[0.29, 0.51]			0.69 (0.07)	[0.55, 0.82]
YBOCS 10	0.61 (0.04)	[0.54, 0.68]			0.60 (0.05)	[0.50, 0.70]
Correlations	Factor Loadings	95% CI				
YBOCS 1-YBOCS 6	0.57 (0.11)	[0.35, 0.79]				
YBOCS 2-YBOCS 7	0.49 (0.11)	[0.28, 0.70]				
YBOCS 3-YBOCS 8	0.01 (0.12)	[-0.21, 0.24]				

Invariance

We evaluated invariance between men and women with our modified model by following the steps outlined by Baldwin (2019). Invariance calculations were performed in Mplus 8 using a WLSMV estimator (see Brown, 2014 for a discussion of the benefits of using this estimator) and scores on the Y-BOCS treated as ordinal data. The first step to calculating invariance is to determine configural invariance. Table 6 display factor loadings by gender for OCD, Obsession, Resistance, and inter-item correlations, respectively. Inspection of these tables indicated that factor loadings were similar between men and women. This pattern of similar loading established configural invariance between men and women (Steenkamp & Baumgartner, 1998). Two goodness-of-fit indices supported good fit with CFI = 0.98 and WRMR = 0.83, one showed adequate fit with RMSEA = 0.08, but one did not with $\chi^2(50) = 100.1$, $p < 0.001$.

Table 6.

Configural Invariance by OCD, Obsessions, and Resistance

	OCD Factor Loadings		Obsessions Factor Loadings		Resistance Factor Loadings	
Items	Men	Women	Men	Women	Men	Women
YBOCS 1	0.95 (0.23)	0.82 (0.20)	1.25 (0.39)	0.95 (0.34)		
YBOCS 2	1.07 (0.19)	0.75 (0.14)	0.95 (0.24)	0.57 (0.20)		
YBOCS 3	1.08 (0.19)	1.10 (0.20)	0.91 (0.20)	0.68 (0.26)		
YBOCS 4	0.53 (0.12)	0.42 (0.11)			0.47 (0.12)	0.57 (0.16)
YBOCS 5	0.61 (0.11)	0.68 (0.13)			0.35 (0.10)	0.54 (0.13)
YBOCS 6	1.20 (0.18)	1.04 (0.15)				
YBOCS 7	2.04 (0.46)	1.71 (0.28)				
YBOCS 8	1.22 (0.17)	1.16 (0.18)				
YBOCS 9	0.87 (0.43)	0.56 (0.13)			1.89 (1.06)	0.82 (0.20)
YBOCS 10	1.15 (0.23)	1.24 (0.28)			1.09 (0.27)	1.19 (0.30)
Inter-Item Correlation	Men	Women				
YBOCS 1- 6	0.62 (0.20)	0.52 (0.13)				
YBOCS 2- 7	0.26 (0.16)	0.68 (0.13)				
YBOCS 3- 8	0.08 (0.18)	-0.12 (0.15)				

Calculating invariance is done by steps to examine the factor structure (configural) and then constrain aspects of the two models in the following order: factor loadings (metric), item intercepts(scalar), and item residuals (residual) to be equal, respectively (Baldwin, 2019). Comparison of the configural and the metric model indicated that the models were not significantly different which suggests that the factor model is metrically invariant, χ^2 (4) = 58.4, p = .867. Comparison of the metric model and the scalar model was not statistically significantly different which also indicated scalar invariance, χ^2 (27) = 19.279, p = .860. The next comparison to consider was the difference between 2 residual models. The first residual model allows for the residual variance to be freely estimated compared to the second residual model that then constrain the residual variance. Comparison of the two residual models was not statistically significantly different which indicated invariance at the residual level, χ^2 (10) = 8.266, p = .603. These last three calculations of measurement invariance indicated that the adjusted Y-BOCS model exhibited invariance between men and women. Table 7 displays fit indices for each step of calculating invariance.

Table 7.

Invariance Fit Indices

Indices	Configural	Metric	Scalar	Residual A	Residual B
χ^2 Model Fit	100.1 (50); $p < .001$	86.1 (64); $p = 0.034$	106.5 (91); $p =.127$	100.0 (81); $p = 0.075$	106.5 (91); $p = .127$
RMSEA	0.08; [0.06, 0.11]	0.05; [0.01, 0.07]	0.34; [0.00, 0.06]	0.04; [0.00, 0.07]	0.34; [0.00, 0.06]
CFI	0.98	0.99	0.99	0.99	0.99
TLI	0.96	0.99	0.99	0.99	0.99
WRMR	0.83	0.90	0.99	0.94	0.99
χ^2 to Baseline	2273.4 (90); $p < .001$	--	--	--	--

We proceeded to calculate structural invariance between genders. This analysis is done in two parts by comparing if the latent factor means and variances between genders. We compared our last residual model to a new model that constrained latent variables variances to be equivalent between groups. Results exhibited invariance at this level due to no statistical difference between the models as indicated by $\chi^2(3) = 2.1$, $p = .54$. There is no statistically significant difference in latent mean variances between genders or no difference in the range of variation of these latent factors between genders. We then compared the constrained variance model to a model with constrained factor means. Results failed to support invariance between factors means due to a significant statistical difference between models as indicated by $\chi^2(3) = 37.1$, $p < .001$. Differences exist between genders on latent factor means. Women scored slightly higher on the general factor of OCD by 0.11 standard deviations but slightly lower on the specific factors of resistance and obsessions by 0.20 and 0.11 standard deviations, respectively.

Convergent Validity

Lastly, we regressed the OCI-R subscales onto the latent variables from our modified CFA model to examine the convergent validity of the latent variables. The results of the analysis are shown in Tables 8 and 9 which display the unstandardized coefficients standardized and intercepts, respectively. Due to missing OCI-R data for some of the participants across sites, the final sample was $N = 260$ for this analysis. OCD significantly predicted Neutralizing and Washing. Obsessions significantly predicted Obsessions, Washing, and Hoarding. Resistance significantly predicted Checking.

Table 8.

Our Model Latent Factors Unstandardized Subscale Beta Coefficients with OCI-R Subscales

	Checking		Hoarding		Neutralizing		Obsessions		Ordering		Washing	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
OCD	-0.01 (0.02)	[-0.06, 0.03]	-0.02 (0.02)	[-0.05, 0.01]	0.05 (0.02)*	[0.01, 0.08]	0.01, (0.02)	[-0.02, 0.04]	0.04 (0.02)	[-0.01, 0.08]	0.03 (0.01)	[0.00, 0.05]
Obsessions	0.03 (0.04)	[-0.05, 0.10]	-0.06 (0.03)	[-0.12, -0.01]	-0.04 (0.03)	[-0.10, 0.02]	0.12 (.03)**	[0.07, 0.17]	-0.04 (0.03)	[-0.10, 0.03]	0.06 (0.02)**	[0.02, 0.09]
Resistance	0.05 (0.02)*	[0.01, 0.09]	-0.01 (0.01)	[-0.04, 0.02]	-0.02 (0.02)	[-0.05, 0.01]	-0.01 (0.01)	[-0.03]	-0.03 (0.02)	[-0.06, 0.00]	0.01 (0.01)	[-0.01, 0.03]
Subscale Mean	5.44 (3.53)		2.99 (3.40)		2.84 (3.44)		6.76 (3.73)		4.62 (3.85)		4.46 (4.23)	

*$p < .05$, **$p < .01$, *** $p < .001$

Table 9.

Our Model Latent Factors Standardized Subscale Beta Coefficients with OCI-R Subscales

	Checking		Hoarding		Neutralizing		Obsessions		Ordering		Washing	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI	β	95% CI
OCD	-0.06 (0.10)	[-0.26, 0.14]	-0.09 (0.07)	[-0.23, 0.05]	0.21 (0.07)**	[0.06, 0.35]	0.03 (0.07)	[-0.12, 0.17]	0.19 (0.10)	[-0.01, 0.39]	0.14 (0.07)*	[0.00, 0.27]
Obsessions	0.08 (0.12)	[-0.15, 0.31]	-0.18 (0.08)*	[-0.34, -0.03]	-0.12 (0.09)	[-0.30, 0.06]	0.39 (0.07)***	[0.26, 0.53]	-0.13 (0.11)	[-0.35, 0.08]	0.21 (0.07)**	[0.08, 0.34]
Resistance	0.31 (0.12)**	[0.08, 0.54]	-0.07 (0.09)	[-0.25, 0.10]	-0.12 (0.10)	[-0.33, 0.07]	-0.06 (0.08)	[-0.22, 0.11]	-0.21 (0.12)	[-0.45, 0.02]	0.07 (0.09)	[-0.10, 0.24]
Subscale Mean	5.44 (3.53)		2.99 (3.40)		2.84 (3.44)		6.76 (3.73)		4.62 (3.85)		4.46 (4.23)	

*$p < .05$, **$p < .01$, *** $p < .001$

Discussion

The main aims of our study were to evaluate the fit of several of the current factor models of the Y-BOCS in the literature on our pooled sample and to evaluate measure and structural invariance using the model exhibited the best fit. Our modified bifactor model exhibited closer fit with our data than the existing models in the literature. The bifactor model also exhibited full measurement invariance between men and women and partial structural invariance. The good fit and gender invariance of our modified model is beneficial for both researchers and clinicians to improve measurement reliability and understanding of OCD symptoms. Despite differences in onset or content of obsessions or compulsions between gender (American Psychiatric Association, 2013; Castle et al.,1995; Labad et al., 2008; Mathis et al., 2011), our invariance analysis indicates that the Y-BOCS measures factor loadings, item intercepts, and item residuals similarly between gender. Results from our convergent validity analysis provided support that our latent factors predicted scores on various OCI-R subscales (except for Ordering).

Several reasons may explain why our bifactor model exhibited higher fit indices than other models. First, one of the main methodological adjustments in our analyses included treating Y-BOCS data as ordinal data rather than interval. As mentioned previously, the anchors for each question vary from item to item. The distribution of responses in Table 4 indicated that most participants responded with the middle of the Likert scale rather than the end points. However, comparison between items is challenging due to the different anchors with each item. Consequently, unless the items are reformatted, the scale should be treated as ordinal data. Second, we allowed for relationships in our model to account for the likely overlap of variance between questions regarding both obsessions and compulsions. For example, the rating of the amount of time spent on obsessions and the amount of time spent engaging in compulsive behaviors is likely related. Third, the poor fit of the Fals-Stewart (1992) model may indicate that while a single global factor for OCD may exist, additional specific factors in the Y-BOCS improve fit. Most of the other models that included multiple factors exhibited better fit indices than the sole global factor model (See Table 5 for review). The best fitting model from the literature was the Deacon model (2005) which included factors of Symptom Severity and Resistance/Control. Consequently, we selected a bifactor model to include a global factor of OCD and two specific factors that included Obsessions and Resistance with correlations allowed between similar items to try to best represent the factor structure of the Y-BOCS and the OCD symptoms it measures. Additionally, the participants sampled at the five sites in the study all were diagnosed with OCD rather than including participants with subclinical levels of OCD which may have increased the power of our study.

Several implications for our findings exist. First, the total OCD scores from the Y-BOCS is still clinically useful. The specific factors did not account for all of the variance. Consequently, the global factor of OCD as measured by the total score remains of clinical value. Intriguingly, despite the original distinction between obsessions and compulsions on the Y-BOCS, no latent factor of compulsion rose from the data. The global factor of OCD may account for the variance associated with compulsions sufficiently and so a specific factor was not needed to account for any additional variance. Alternatively, it is possible that the specific factor of Resistance subsumed the variance associated with compulsions. The specific factor of Resistance included the items regarding resistance against and degree of control over obsessions and compulsions. These two aspects of compulsions may be the most salient for those experiencing compulsions rather than time, interference or distress associated with compulsions.

Additionally, our factors predicted scores on all of the OCI-R subscales, except for Ordering. The Ordering subscale of the OCI-R includes questions arranging objects. Somewhat surprisingly, neither the global OCD factor nor the specific Resistance factor significantly predicted the Ordering subscale. The specific Obsessions factor predicted multiple OCI-R subscales with the strongest relationship with the Obsessions subscale which supports the validity of our factor. The relationship between Resistance and Checking was positive which seemed counterintuitive at first. If a client is reporting higher resistance to OCD symptoms, then wouldn't he or she also check less frequently? However, perhaps those who report a high level of resistance engage in the checking behavior to relieve the anxiety associated with that resistance while those with low resistance to those symptoms feel little need to check behaviors.

Replication of our model structure in other samples is needed. Possible variables that could modify the models include time from diagnosis, level of patient insight, content themes of obsessions and compulsions, and patient age. For example, since the Y-BOCS is sensitive to drug administration during the course of treatment (Goodman et al., 1989a, b), sampling patients at varying times from onset of symptoms would add to the literature. One of the limitations of our study was that time from diagnosis was not collected at each of the five sample sites. If our model is not supported at different stages from the onset of symptoms, then perhaps different latent factors affect responses to the Y-BOCS at different stages of treatment for OCD. Consequently, adding other items to measure potential new factors would be needed. Additionally, the latent factor Resistance may indirectly measure the level of patient insight. Poorer client insight is often associated with poorer outcomes (American Psychiatric Association, 2013). Consequently, if the Y-BOCS could help to assess level of insight more objectively, then it may benefit clinicians to tailor treatment with their clients more appropriately. Perhaps including differing degrees of client insight as a variable would provide additional evidence of the importance of Resistance in the factor model in the Y-BOCS and determine the level of similarity between client insight and Resistance. Additionally, our sample was not large enough to split to confirm our factor model.

In conclusion, we offer a factor model of the Y-BOCS that exhibited full measurement and partial structural invariance between genders. We recommend that other researchers treat Y-BOCS data as ordinal rather than continuous data. Replication of our factor structure in a variety of samples will be needed to provide further support for the validity of this structure. Further research is needed to evaluate the invariance utilizing other comparison groups such as age or temporal invariance across the course of OCD.

References

Abramowitz, J. S., & Deacon, B. J. (2006). Psychometric properties and construct validity of the Obsessive–Compulsive Inventory—Revised: Replication and extension with a clinical sample. *Journal of Anxiety Disorders*, *20*(8), 1016-1035.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Amir, N., Foa, E. B., & Coles, M. E. (1997). Factor structure of the Yale-Brown Obsessive Compulsive Scale. *Psychological Assessment, 9,* 312-316.

Anholt, G. E., van Oppen, P., Cath, D. C., Smit, J. H., den Boer, J. A., Verbraak, M. J. P. M., & van Balkom, J. L. M. (2010). The Yale-Brown Obsessive Compulsive Scale: Factor structure of a large sample. *Frontiers in Psychiatry*, *1*, 1-7.

Antony, M. M., Orsillo, S. M., & Roemer, L. (2001). *Practitioner's guide to empirically based measures of anxiety*. New York, NY: Kluwer Academic/Plenum Publishing.

Arrindell, W. A., de Vlaming, I. H., Eisenhardt, B. M., van Berkum, D. E., & Kwee, M. G. T. (2002). Cross-cultural validity of the Yale-Brown Obsessive Compulsive Scale. *Journal of Behavior Therapy and Experimental Psychiatry, 33,* 159-176.

Baglin, J. (2014). Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research & Evaluation*, *19*(5), 2.

Baldwin, S. (2019). *Psychological statistics and psychometrics using Stata.* College Station, TX: Stata Press.

Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, *50*(2), 1-5.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research*. Guilford Publications.

Brown, T. A., Barlow, D. H., & DiNardo, P. A. (1994). *Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV): Client interview schedule*. Graywind Publications Incorporated.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions, 154*, 136-136.

Castle, D. J., Deale, A., & Marks, I. M. (1995). Gender differences in Obsessive Compulsive Disorder. *Australian & New Zealand Journal of Psychiatry, 29*(1), 114-117.

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, *80*(1), 219-251.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7), 1-9.

Deacon, B. J. & Abramowitz, J. S. (2005). The Yale-Brown Obsessive Compulsive Scale: Factor analysis, construct validity, and suggestions for refinement. *Journal of Anxiety Disorders, 19,* 573-585.

DiStefano, C., Liu, J., Jiang, N., & Shi, D. (2018). Examination of the weighted root mean square residual: Evidence for trustworthiness? *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 453-466.

Fals-Stewart, W. (1992). A dimensional analysis of the Yale-Brown Obsessive Compulsive Scale. *Psychological Report, 70,* 239-240.

First, M.B. & Gibbon, M. (2004). The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-1) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). In M. J. Hilsentroth & D. L. Segal (Eds), *Comprehensive handbook of psychological assessment, Vol. 2 Personality assessment* (pp. 134-143). Hoboken, NJ: John Wiley.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485.

Forrest, M., & Andersen, B. (1986). Ordinal scale and statistics in medical research. *British Medical Journal (Clinical Research Ed.), 292*(6519), 537-538.

Garnaat, S. L. & Norton, P. J. (2010). Factor structure and measurement invariance of the Yale-Brown Obsessive Compulsive Scale across four racial/ethnic groups. *Journal of Anxiety Disorders, 24,* 723-728.

Gibbons, R. D., & Hedeker, D. R. (1992). Full-information item bi-factor analysis. *Psychometrika*, *57*(3), 423-436.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., ... & Charney, D. S. (1989a). The Yale-Brown Obsessive Compulsive Scale: I. Development, use, and reliability. *Archives of General Psychiatry*, *46*(11), 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*(11), 1012-1016.

Hajcak, G., Huppert, J. D., Simons, R. F., & Foa, E. B. (2004). Psychometric properties of the OCI-R in a college sample. *Behaviour Research and Therapy*, *42*(1), 115-123.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*(3), 117-144.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Huppert, J. D., Walther, M. R., Hajcak, G., Yadin, E., Foa, E. B., Simpson, H. B., & Liebowitz, M. R. (2007). The OCI-R: Validation of the subscales in a clinical sample. *Journal of Anxiety Disorders*, *21*(3), 394-406.

Kim, S. W., Dysken, M. W., Pheley, A. M., & Hoover, K. M. (1994). The Yale-Brown Obsessive Compulsive Scale: Measures of internal consistency. *Psychiatry Research, 51,* 203-211.

Labad, J., Menchon, J. M., Alonso, P., Segalas, C., Jimenez, S., Jaurrieta, N., ... & Vallejo, J. (2008). Gender differences in obsessive–compulsive symptom dimensions. *Depression and Anxiety*, *25*(10), 832-838.

Lance, C. E., & Vandenberg, R. J. (2002). Confirmatory factor analysis. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 221-254). San Francisco, CA, US: Jossey-Bass.

Larson, M.J. 2012. [Cognitive Control Disruption in OCD]. Unpublished raw data.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936-949.

López-Pina, J. A., Sánchez-Meca, J., López-López, J. A., Marín-Martínez, F., Núñez-Núñez, R. M., Rosa-Alcázar, A. I., ... & Ferrer-Requena, J. (2015). The Yale–Brown Obsessive Compulsive Scale: A reliability generalization meta-analysis. *Assessment*, *22*(5), 619-628.

Mathis, M. A. D., Alvarenga, P. D., Funaro, G., Torresan, R. C., Moraes, I., Torres, A. R., ... & Hounie, A. G. (2011). Gender differences in Obsessive-Compulsive Disorder: A literature review. *Brazilian Journal of Psychiatry*, *33*(4), 390-399.

McKay, D., Danyko, S. Neizroglu, F., & Yaryura-Tobias, J. A. (1995). Factor structure of the Yale-Brown Obsessive-Compulsive Scale: A two dimensional measure. *Behaviour Research and Therapy, 33,* 865-869.

McKay, D., Neziroglu, F., Stevens, K., & Yaryura-Tobias, J. A. (1998). The Yale-Brown Obsessive-Compulsive Scale: Confirmatory factor analytic findings. *Journal of Psychopathology and Behavioral Assessment*, *20*(3), 265-274.

Moritz, S., Meier, B., Kloss, M. Jacobsen, D., Wein, C., Fricke, S., & Hand, I. (2002). Dimensional structure of the Yale-Brown Obsessive Compulsive Scale (Y-BOCS). *Psychiatry Research, 109,* 193-199.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*(3), 552.

Rosenfeld, R., Dar, R., Anderson, D., Kobak, K. A., & Greist, J. H. (1992). A computer-administered version of the Yale-Brown Obsessive-Compulsive Scale. *Psychological Assessment*, *4*(3), 329.

Steenkamp, J. B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78-90.

Stevens, S. S. (1946). On the theory of scales of measurement.

Storch, E. A., Abramowitz, J. S., & Keeley, M. (2009). Correlates and mediators of functional disability in Obsessive–Compulsive Disorder. *Depression and Anxiety*, *26*(9), 806-813.

Storch, E. A., Larson, M. J., Shapira, N. A., Ward, H. E., Murphy, T. K., Geffken, G. R., ... & Goodman, W. K. (2006). Clinical predictors of early fluoxetine treatment response in Obsessive–Compulsive Disorder. *Depression and Anxiety*, *23*(7), 429-433.

Storch, E. A., Merlo, L. J., Bengtson, M., Murphy, T. K., Lewis, M. H., Yang, M. C., ... & Geffken, G. R. (2007). D-cycloserine does not enhance exposure–response prevention therapy in Obsessive–Compulsive Disorder. *International Clinical Psychopharmacology*, *22*(4), 230-237.

Storch, E. A., Merlo, L. J., Lehmkuhl, H., Geffken, G. R., Jacob, M., Ricketts, E., ... & Goodman, W. K. (2008). Cognitive-behavioral therapy for Obsessive–Compulsive Disorder: A non-randomized comparison of intensive and weekly approaches. *Journal of Anxiety Disorders*, *22*(7), 1146-1158.

Storch, E. A., Shapira, N. A., Dimoulas, E., Geffken, G. R., Murphy, T. K., & Goodman, W. K. (2005). Yale-Brown Obsessive Compulsive Scale: The dimension structure revisited. *Depression and Anxiety, 22,* 28-35.

Taylor, S. (1995). Assessment of obsessions and compulsions: Reliability, validity, and sensitivity to treatment effects. *Clinical Psychology Review*, *15*(4), 261-296.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research*, 281-324.

Woody, S. R., Steketee, G., & Chambless, D. L. (1995). Reliability and validity of the Yale-Brown Obsessive-Compulsive Scale. *Behaviour Research and Therapy*, *33*(5), 597-605.