Ethnic Identification Biases Responses to the Padua Inventory for Obsessive-Compulsive Disorder

Monnica T. Williams
Eric Turkheimer
Karen M. Schmidt
University of Virginia, Charlottesville

Thomas F. Oltmanns
Washington University

The authors report differential item functioning (DIF) between Black and White participants completing the 60-item Padua Inventory (PI) for obsessive-compulsive disorder (OCD). The authors use an Internet-generated sample that included 105 Blacks, 67 Hispanics, 582 Whites, and 136 additional participants reporting an OCD diagnosis. Factor analysis replicated prior work indicating the PI consists of four factors: contamination fears, checking behaviors, impaired control over thoughts, and fear of losing control over impulses. On the contamination subscale, nonclinical Black and Hispanic mean scores were as high as the OCD group. Comparing Blacks to Whites, the authors applied an item response theory, DIF-graded response model to each factor and found significant DIF on eight items, with biased items in each factor. Results suggest that extraneous factors contribute to racial differences on scores. Cultural practices and fear of being negatively stereotyped may contribute to item bias.

Keywords: obsessive-compulsive disorder; item response theory; ethnic differences; assessment

OBSESSIVE-COMPULSIVE DISORDER (OCD) INSTRUMENTS

OCD is often identified by mental health professionals and researchers through the use of screening tools, such as questionnaires or checklists that ask patients about obsessive and compulsive symptoms. For U.S. minority groups, such as African Americans, assessment techniques for OCD have not been the subject of much scientific inquiry. Only three studies have examined racial differences in the prevalence of OCD, and the results are inconsistent. The most extensive data regarding this issue come from the Epidemiologic Catchment Area (ECA) project, which reported that the lifetime prevalence of OCD among Blacks was significantly lower than among Whites (Karno, Golding, Sorenson, & Burnam, 1988). Two smaller studies found differences in the opposite direction (Faberga, Mezzich, & Ulrich, 1988; Valleni-Basile et al., 1996). Several methodological factors may account for these discrepancies, including the investigators' choice of instruments and the fact that there has been little research done on the validity of these instruments for minority groups.

The Maudsley Obsessional Compulsive Inventory (MOCI; Hodgson & Rachman, 1977) was found to lack

predictive validity for African Americans when administered to a large sample of college students, in part because of overendorsement of cleaning and checking items (Thomas, Turkheimer, & Oltmanns, 2000). The 13-item National Anxiety Disorders Screening Day (NADSD) instrument was assessed based on data from a large sample of patients from five ethnic groups (Ritsher, Stuening, Hellman, & Guardino, 2002). A six-factor model fit the data very well for Whites, but some areas of the model did not fit as well for certain ethnic groups. For example, within the three-item OCD scale, the question about compulsions was problematic because it loaded on a different factor for Blacks, as did some items on Post-Traumatic Stress Disorder and Generalized Anxiety Disorder scales. Blacks were also more likely to report OCD symptoms in this study.

THE PADUA INVENTORY (PI)

The goal of this study was to examine the psychometric properties of another important OCD questionnaire, the PI (Sanavio, 1988), and to examine some hypotheses for observed racial differences in item functioning. Sanavio (1988) reported that the PI is reliably correlated with other measures of OCD and also discriminates between participants with OCD versus related conditions, such as other anxiety disorders. The inventory consists of 60 polytomously scored items, each with five response levels. Participants rate each question based on the level of distress it produces ($0 = not\ at\ all$, $4 = very\ much$). The PI evaluates a person's tendency to worry and doubt (obsessions) and perform behaviors intended to ward off those doubts (compulsions), in four main areas: (a) contamination, (b) checking, (c) impaired control over mental activities, and (d) worries about losing control over one's behaviors. The test is scored by summing the responses to each of the 60 questions.

ASSESSING BIAS

Although correlations among tests and factor analysis can provide some information about scale equivalence, item response theory offers a method for examining equivalence by studying psychometric item bias, or differential item functioning (DIF). Each item in a unidimensional scale is analyzed to determine how likely participants in two groups (e.g., Black and White participants in the current study) are to endorse a response category, conditional on their score on the latent trait being measured by the scale. DIF is defined as the unexpected difference in response to a test item between two groups, once the attribute the test is measuring (in this case an OCD-related anxiety trait) is held constant. Although DIF traditionally refers to items in educational or intelligence tests, the concept can also apply to other psychological measures. This approach has been used recently in the cross-cultural analysis and foreign language equivalence of psychological measures for depression, psychopathy, and posttraumatic stress (Azocar, Arean, Miranda, & Munoz 2001; Cooke, Kosson, & Michie, 2001; Orlando & Marshall, 2002).

METHOD

Participants

Information was gathered from participants using an online multiple-choice questionnaire, which was presented when individuals followed links to the survey from announcements placed on Internet bulletin boards, electronic mailing lists, OCD-related Web sites, health Web sites, and popular Internet search engines. Participants were invited to complete a survey asking about their propensity to worry or doubt regarding everyday concerns. They were not required to provide names or any other identifying information.

Data were stored offline to preserve security of the data; risk to participants must therefore be considered minimal. Although reactive effects were not specifically addressed, participants were able to request a personalized response if they required further information. Participants were informed that data provided would be used for scientific research and statistical analyses and that they could cease participation at any time. Once completed, respondents were offered links to more information about OCD.

There were 1,083[1] participants, with a mean age of 23.3 ($SD = 10.2$). For the purpose of this analysis, participants were divided into five groups:

- Whites/Caucasians not reporting OCD ($n = 582$),
- Blacks/African Americans not reporting OCD ($n = 105$),
- Hispanics not reporting OCD ($n = 67$),
- persons reporting possible OCD symptoms (primarily Whites but all races included; $n = 175$), and
- persons reporting a prior OCD diagnosis (primarily Whites but all races included; $n = 136$).

Possible OCD was determined by asking respondents if they had ever thought or been told by others that they had OCD, in the absence of a formal diagnosis. The OCD diagnosis group consists of those who had been told by a health professional that they had OCD. Those selecting Asian, Middle Eastern, or other category without possible/proba-

ble OCD were eliminated from the sample, as were persons who left more than seven items blank (252 participants). Participants were 26.5% male and 70.7% female. The vast majority of participants were from the United States (89.2%), and the remainder were primarily from English-speaking countries (e.g., Canada, United Kingdom, Australia). The mean age was 23.3.

Measures

In addition to demographic and mental health history questions, participants were asked to complete the PI. The 5-item rating scale used for the PI, also used for the shorter Washington State University revised version (PI-WSUR; Burns, Keortge, Formea, & Sternberger, 1996), employs the following wording for categories, based on the amount of distress caused by each item:

- 0 = *not at all*
- 1 = *a little*
- 2 = *quite a lot*
- 3 = *a lot*
- 4 = *very much*

Because we were concerned that the wording of the middle category might be misunderstood in light of the terms used for other categories intended to represent greater distress, we replaced "quite a lot" with "somewhat." Participants were able to leave questions blank if desired and were given room to make their own free-response comments.

Statistical Procedures

To verify the factor structure of the PI, we performed a varimax-rotated exploratory factor analysis (EFA) using M-PLUS (Muthén & Muthén, 1998). Varimax rotation was selected to facilitate comparison with previous factor analyses. We elected to conduct an EFA rather than a series of confirmatory factor analyses (CFAs) of previously reported results, because although the results of previous studies were generally quite comparable, they differed in a number of small details at the item level. Our goal was to show general conformity with previous results and to provide a basis for our subsequent item response theory analysis rather than select among the small differences in the previous factor solutions. M-PLUS is a general factor analysis and structural equation modeling program with special facilities for modeling categorical data. The M-PLUS factor analysis model views the categorical items as arising from latent continuous variables that are divided into k categories at $k-1$ thresholds on an underlying latent normal distribution. The exploratory factor analysis is conducted on the covariances among the latent continuous variables.

The DIF analysis was performed with PARSCALE 3.2 using the Modified Graded Response Model (M-GRM; Muraki, 1990), designed for use with polytomous data. In this model, each item has a slope, α_j, and a difficulty, b_j. (The term *difficulty* is borrowed from the ability domain, where IRT models originated. It refers to the overall probability that a participant will endorse an item, or "get it right." In personality and psychopathology scales, the difficulty parameter refers to likelihood of a pathological response to an item, for a fixed level of trait in the respondent. Difficult items are those that only participants with high values on the trait in question are likely to endorse.) The thresholds for the response categories on each item are represented by c_k. Each participant has a latent score on the trait, denoted as θ_i. The probability that participant i will endorse item j using response category k is given by,

$$P(y_{ijk} = 1) = \frac{\exp[\alpha_j(\theta_i - b_j + c_k)]}{1 + \exp[\alpha_j(\theta_i - b_j + c_k)]}. \qquad (1.1)$$

Equation 1.1 appears complex, but the model is actually quite simple. If the likelihood of an item endorsement is expressed as a log odds rather than a probability, it is linear:

$$\log \text{odds}(y_{ijk} = 1) = \theta_i - b_j + c_k. \qquad (1.2)$$

In general, the model expresses the expectation that the log of the odds of an endorsement response increases linearly with the amount of the trait possessed by the respondent and decreases linearly with the extremity of the item and with increasing points on the response scale.

Because there were many questions (60), several categories of questions (5), and a relatively small sample size for some groupings, the PARSCALE software was limited at times in its ability to perform the analysis. For the IRT analysis, it was necessary to collapse response categories (0,1,2,3,4 became 1,2,2,3,3, respectively) to simplify the estimation.

The DIF analysis was conducted in PARSCALE by fitting the M-GRM model separately in the racial groups for each of the factors identified by the exploratory factor analysis, allowing the mean value of the scale to differ between the groups but constraining the solutions so the slopes and category thresholds of the items were equal across groups. The item difficulties were free to differ between the groups but were scaled so they summed to zero in each group. The group differences between the item difficulties were then calculated and divided by their standard

TABLE 1
Mean Padua Scores

Study	n	Sample	M	(SD)
Nonclinical samples				
Sanavio (1988)	1,357	Italian community	53.6	(27.7) male
			62.5	(29.2) female
Sternberger and Burns (1991)	678	American students	41.3	(25.7)
Van Oppen (1992)	430	Dutch community	25.8	(20.8) male
			29.4	(20.7) female
Burns, Formea, Keortge, and Sternberger (1995)	2,287	American students	35.9	(30.0)
Kyrios, Bhar, and Wade (1996)	306	Australian students	42.7	(26.4)
This study	754	Internet participants	38.3	(30.3) White
			42.9	(33.3) Black
			47.8	(33.9) Hispanic
Clinical samples				
Sanavio (1988)	75	Italian OCD patients	83.6	(34.8) male
			98.6	(32.3) female
Hafner (1988)	81	Australian self-help group	66.9	(33.7)
Kyrios and Iob (1998)	15	Australian outpatients	73.27	(39.34)
This study	136	Internet OCD participants	83.5	(43.5)

NOTE: OCD = obsessive-compulsive disorder.

error. This procedure has the effect of identifying items with difficulty parameters that differ between the two groups after controlling for the average difference between the two groups. For example, if one group endorses most items on a trait more frequently than the other group, but there is a small group of items for which the pattern is reversed, these items will be identified as biased by the DIF procedure.

For example, suppose a washing scale has been estimated as described above in White and Black participants and, for simplicity, that the results have been scaled in terms of log odds rather than the traditional probabilities. Blacks are (hypothetically) more likely to endorse items on the scale, resulting in an average score for Blacks on the latent ability of 0.41, indicating that the log of the odds that a Black participant will endorse the item is 0.41 units greater than the log of the odds a White participant will endorse it. (The odds ratio is thus $e^{.41} = 1.5$.) Now, consider a particular item, "I always shower after touching domestic animals," with an a parameter equal to 0.6 in both groups. According to the model, the predicted difference between Blacks to Whites in the log odds of endorsement for the item is equal to $(0.41)(0.6) = 0.25$. However, on examining the results, it turns out that Blacks are even more likely to endorse this item relative to Whites than is true on the scale in general, with a difference in log odds equal to 0.8. The difference between the observed value of 0.8 and the value of 0.25 predicted by the no-bias model can be compared to its standard error and tested for significance. In this hypothetical example, we would conclude that differences in the probability of endorsing the animal item cannot be explained by the overall group difference in the washing scale.

To the extent the property does hold for all items, it indicates that group differences at the item level can be completely accounted for by a single group difference in the latent trait. To the extent some items do not fit this model, it suggests that responses are influenced by group membership in addition to participants' scores on the latent trait, which is a definition of DIF. An important implication of this procedure is that it separates overall group differences on a trait from item-level group differences after differences at the level of the trait have been controlled. The latter is indicative of DIF, but the former is not.

RESULTS

Preliminary Analysis

As shown in Table 1, mean scores on the PI for both majority and minority members were consistent with earlier findings for nonclinical U.S. samples (Burns, Formea, Keortge, & Sternberger, 1995; Sternberger & Burns, 1991). No significant differences in overall scores emerged between Whites, with a mean of 38.4 ($SD = 30.3$), and Blacks, with a mean of 42.9 ($SD = 33.3$), $t(686) = -1.38$, $p = .17$, two-tailed test. Hispanics had a mean score of 47.8 ($SD = 33.9$), which did differ significantly from Whites, $t(648) = -2.36$, $p = .02$, two-tailed test. Although a low N prevents further analysis at the item level, mean scores on the PI are provided for this group, because to our knowledge, these have not been previously reported for Hispanic respondents.

Mean scores for persons with possible and probable OCD were 69.6 ($SD = 41.4$) and 83.5 ($SD = 43.5$), respec-

tively, and differed in the expected direction. This is consistent with prior studies done with clinical samples (Hafner, 1988; Kyrios & Iob, 1998; Sanavio, 1988).

Factor Analysis

Exploratory factor analysis was performed on the entire sample, omitting any participants who left one or more responses blank, for a total of 785. Four factors were selected for rotation based on examination of scree plots and to simplify comparison with earlier studies, all of which had identified four factors in the PI (Kyrios, Bhar, & Wade, 1996; Sanavio, 1988; Sternberger & Burns, 1991). The data most closely match those of Sternberger and Burns (1991), who sampled U.S. college students in Washington State. The rotated factors were identified as (a) contamination and washing, (b) checking, (c) doubting and loss of control over mental activities, and (d) fear of acting on impulses. Table 2 details these findings.

Mean Raw Scores by Factor

Mean raw scores for each group, shown in Table 3, were examined separately by factor. For the contamination items, Whites scored significantly lower than persons reporting an OCD diagnosis. However, mean scores for Blacks and Hispanics were significantly higher than scores for Whites and did not differ significantly from the OCD group. An ANOVA found no significant mean differences between ethnic groups for the remaining three factors: Checking, Mental Control/Doubt, and Fear of Impulses. As expected, OCD patients scored significantly higher on each of these subscales than any did nonaffected group.

DIF: Four Factors

The DIF analysis was performed for each of the factors for Blacks and Whites. As mentioned previously, Hispanics were not included in the IRT analysis due to a low n. Results appear in Table 4, which lists raw score means and standard deviations, b (item difficulty) values, standard errors, differences between b values, and the critical ratio of group differences to their standard error. Items with critical ratios ($b_2 - b_1/\sqrt{SE^2_2 + SE^2_1}$) greater than 1.96 (or less than –1.96) were considered significant and are indicated with a superscripted a. Items with negative differences indicate items more frequently endorsed by Blacks, whereas items with positive differences were more frequently endorsed by Whites. Some items that loaded on more than one factor appear twice.

Note, as described above, that the DIF models for the items are independent of any group differences on the entire scale. So, for example, on Factor 1 (Contamination), Blacks scored higher than Whites by 0.44 units on the IRT theta scale.[2] Again as above, the scale is essentially the log of an odds ratio, so the odds Blacks would endorse an item on this scale were $e^{.44}$ or 1.55 times higher than the equivalent odds for Whites. The two items identified with DIF on this scale deviate from this pattern. Whites were more likely to report that their hands felt dirty when touching money, whereas Blacks were much more likely to report feeling dirty after touching an animal. There was no overall mean difference on either the checking or the mental control/doubt subscales, yet four checking items and two mental control items were endorsed significantly more frequently by Blacks. On the fear of impulses subscale, Whites were more likely than Blacks to report impulses to tear off their clothes in public, whereas Blacks were more likely to endorse an item concerning unintended obscene thoughts. On the mental control subscale, Blacks were more likely to endorse an item about unwanted unpleasant thoughts, although this item was only marginally significant.

Figure 1 is a graph of the probability that Black and White participants will endorse Item 10 ("If an animal touches me, I feel dirty and immediately have to wash myself or change my clothing"), plotted as a function of estimated score on the latent contamination scale. Graphs of this kind are called *item characteristic curves* (ICC). The figure illustrates that Blacks are more likely than Whites to endorse the item, almost 20% more likely in the middle range of the latent trait, even after the mean difference between Blacks and Whites has been accounted for. Therefore, simple scoring procedures based on summing the number of endorsed items will tend to overestimate the trait in Blacks relative to Whites.

DISCUSSION

Reasons for Differences

Many of the DIF findings are consistent with earlier work in examining differences in mean item scores between Blacks and Whites on the MOCI (Thomas et al., 2000). That study found overreporting of cleaning and grooming behaviors in Blacks compared to Whites, as did this investigation. There are several possible reasons why Blacks and Whites may differ in their response to OCD questionnaires. Some of these differences may be representative of different cultural practices and norms regarding normal behaviors (washing, grooming, being cautious), related to the pathological behaviors being assessed. Other differences may relate to how underrepresented minorities perceive majority observers. For

TABLE 2
Factor Structure of the Padua Inventory

Item	Contamination				Checking				Mental Control/Doubt				Fear of Impulses			
	Sanavio	Kyrios	Sternberger	Williams	Sanavio	Kyrios	Sternberger	Williams	Sanavio	Kyrios	Sternberger	Williams	Sanavio	Kyrios	Sternberger	Williams
1	0.52		0.43	0.54												
2	0.58	0.64	0.50	0.58												
3	0.64	0.62	0.63	0.66												
4	0.57	0.65	0.63	0.58												
5	0.61	0.69	0.63	0.60												
6	0.56	0.46	0.66	0.65												
7	0.66	0.53	0.61	0.66												
8	0.67	0.61	0.51	0.66												
9	0.65	0.64	0.52	0.66												
10	0.64	0.65	0.59	0.50												
11									0.44		0.41	0.53				
12						0.44					0.53	0.55				
13						0.50					0.40	0.53				
14		0.56						0.43								
15		0.55						0.54								
16		0.50						0.45								
17								0.47								
18					0.49	0.51	0.44	0.67			0.40					
19					0.63	0.64	0.64	0.73								
20		0.68			0.65		0.71	0.73								
21					0.69	0.57	0.68	0.65								
22		0.64			0.65		0.69	0.63								
23					0.52	0.55	0.67	0.64								
24					0.52	0.44	0.62	0.47								
25					0.59	0.61	0.65	0.51								
26						0.42			0.51		0.61	0.62				
27						0.51		0.517	0.49		0.55	0.55				
28						0.48			0.51		0.68	0.72				
29						0.48		0.413	0.49	0.45	0.68	0.66				
30						0.50		0.521	0.42			0.42				
31						0.44			0.65	0.43	0.65	0.68				
32									0.72	0.47		0.63				
33									0.68	0.50	0.64	0.69				
34									0.44							0.47
35									0.54	0.47	0.53	0.61				0.40
36									0.63	0.55	0.55	0.56				
37						0.42			0.50	0.54	0.52	0.49				
38									0.47	0.73						0.40
39										0.47		0.52				
40								0.45		0.56						
41								0.49		0.46						
42						0.60		0.40			0.50	0.47				
43						0.46		0.46	0.46	0.44	0.45	0.42				
44									0.43		0.46	0.60				
45											0.53	0.62				
46													0.57	0.62	0.52	0.62
47													0.48	0.60	0.57	0.61
48															0.48	0.59
49													0.42	0.67	0.66	0.53
50														0.56	0.67	0.67
51																
52										0.43		0.48		0.51	0.47	0.48
53													0.62	0.62	0.69	0.69
54													0.47	0.59	0.62	0.61
55													0.44	0.54	0.52	0.62
56														0.52	0.61	0.59
57													0.48	0.45		0.42
58										0.43	0.42					
59									0.45	0.41	0.48	0.46				
60	0.43	0.41														

NOTE: Data are from Sanavio (1988); Kyrios, Bhar, and Wade (1996); Sternberger and Burns (1991); and this study (Williams, Turkheimer, Schmidt, and Oltmanns).

TABLE 3
Mean Subscores by Group

Group	n	Factor 1 Contamination		Factor 2 Checking		Factor 3 Mental Control/Doubt		Factor 4 Fear of Impulses		Total
		M	SD	M	SD	M	SD	M	SD	
White	582	7.91	6.21	10.94	10.77	17.41	15.61	5.86	7.48	38.28
Black	105	10.77	7.28	12.42	11.97	18.59	17.60	5.88	8.57	42.88
Hispanic	67	10.52	6.72	13.64	11.99	20.48	17.11	7.81	8.08	47.76
OCD-sx	175	11.18	8.32	23.47	17.30	32.55	19.85	10.37	10.61	41.40
OCD-dx	136	11.55	9.49	25.97	17.73	41.82	20.47	14.71	11.22	83.49

OCD = obsessive-compulsive disorder; OCD-sx = people who believe they may have OCD; OCD-dx = people diagnosed by a medical professional as having OCD.

TABLE 4

Item	Group	Raw Score M	SD	b (location)	SE	DIF: $b_2 - b_1$	Critical Ratio	Question
Factor 1: Contamination								
1[a]	1: White	0.79	0.99	1.417	0.080	0.472	2.310	I feel my hands are dirty when I touch money.
	2: Black	0.87	1.19	1.889	0.188			
2	1: White	0.69	0.92	1.440	0.072	0.206	1.144	I think even slight contact with bodily secretions . . . may contaminate my clothes or somehow harm me.
	2: Black	0.86	1.05	1.646	0.165			
3	1: White	0.59	0.81	1.537	0.061	0.005	0.033	I find it difficult to touch an object when I know it has been touched by strangers or by certain people.
	2: Black	0.85	0.96	1.542	0.138			
4	1: White	1.42	1.14	0.099	0.065	0.024	0.144	I find it difficult to touch garbage or dirty things.
	2: Black	1.78	1.31	0.123	0.153			
5	1: White	1.02	1.09	0.802	0.064	0.096	0.592	I avoid using public toilets because I am afraid of disease and contamination.
	2: Black	1.29	1.22	0.898	0.149			
6	1: White	0.54	0.85	1.832	0.070	0.264	1.544	I avoid using public telephones because I am afraid of contagion and disease.
	2: Black	0.66	0.93	2.096	0.156			
7	1: White	0.58	0.88	1.762	0.075	−0.050	−0.280	I wash my hands more often and longer than necessary.
	2: Black	0.87	1.13	1.712	0.162			
8	1: White	0.62	0.92	1.695	0.074	0.022	0.123	I sometimes have to wash or clean myself simply because I think I may be dirty or "contaminated."
	2: Black	0.89	1.17	1.717	0.163			
9	1: White	1.21	1.22	0.590	0.074	−0.278	−1.485	If I touch something I think is "contaminated," I immediately have to wash or clean myself.
	2: Black	1.75	1.51	0.312	0.172			
10[a]	1: White	0.45	0.81	2.239	0.104	−0.759	−3.490	If an animal touches me, I feel dirty and immediately have to wash myself or change my clothing.
	2: Black	1.05	1.29	1.480	0.191			
Factor 2: Checking								
14	1: White	0.71	1.04	1.424	0.089	0.191	0.862	I feel obliged to follow a particular order in dressing, undressing, and washing myself.
	2: Black	0.75	1.20	1.615	0.203			
15	1: White	0.62	1.00	1.655	0.095	0.010	0.044	Before going to sleep, I have to do certain things in a certain order.
	2: Black	0.68	1.12	1.665	0.208			
16	1: White	0.20	0.59	0.000	0.372	0.005	0.005	Before going to bed, I have to hang up or fold my clothes in a special way.
	2: Black	0.17	0.60	0.005	0.878			
17	1: White	0.24	0.74	3.277	0.196	0.503	1.155	I feel I have to repeat certain numbers for no reason.
	2: Black	0.18	0.69	3.780	0.389			
18	1: White	0.53	0.87	1.536	0.065	0.221	1.322	I have to do things several times before I think they are properly done.
	2: Black	0.51	0.95	1.757	0.154			

(continued)

TABLE 4 (continued)

Item	Group	Raw Score M	SD	b (location)	SE	DIF: $b_2 - b_1$	Critical Ratio	Question
19	1: White	0.74	1.00	1.072	0.056	0.144	0.992	I tend to keep on checking things more often than necessary.
	2: Black	0.64	0.94	1.216	0.134			
20	1: White	0.32	0.73	2.168	0.090	−0.064	−0.313	I check and recheck gas and water taps and light switches after turning them off.
	2: Black	0.35	0.70	2.104	0.184			
21[a]	1: White	0.26	0.66	2.661	0.131	−0.529	−2.108	I return home to check doors, windows, drawers, etc., to make sure they are properly shut.
	2: Black	0.47	1.00	2.132	0.214			
22	1: White	0.76	0.94	0.948	0.063	−0.251	−1.606	I keep on checking forms, documents, checks, etc. in detail to make sure I have filled them in correctly.
	2: Black	1.07	1.25	0.697	0.143			
23	1: White	0.50	0.90	1.821	0.091	0.235	1.048	I keep on going back to see that matches, cigarettes, etc. are properly extinguished.
	2: Black	0.49	0.96	2.056	0.205			
24	1: White	0.98	1.04	0.632	0.057	-0.124	-0.852	When I handle money, I count and recount it several times.
	2: Black	1.15	1.32	0.508	0.134			
25[a]	1: White	0.77	1.00	1.033	0.065	−0.375	−2.260	I check letters carefully many times before posting them.
	2: Black	1.12	1.24	0.658	0.145			
27	1: White	0.93	1.06	0.750	0.053	−0.035	−0.258	Sometimes I am not sure I have done things which in fact I know I have done.
	2: Black	1.02	1.10	0.715	0.125			
29	1: White	0.89	1.07	0.874	0.059	0.008	0.053	After doing something carefully, I still have the impression I have either done it badly or not finished it.
	2: Black	0.90	1.14	0.882	0.138			
30[a]	1: White	0.46	0.89	1.914	0.086	−0.495	−2.686	I am sometimes late because I keep on doing certain things more often than necessary.
	2: Black	0.73	1.20	1.419	0.163			
40	1: White	0.54	0.96	1.876	0.101	0.259	1.046	I sometimes start counting objects for no reason.
	2: Black	0.47	0.91	2.135	0.226			
41	1: White	0.34	0.81	2.458	0.123	0.563	1.835	I feel I have to remember completely unimportant numbers.
	2: Black	0.24	0.72	3.021	0.281			
42[a]	1: White	0.87	1.08	0.890	0.063	−0.346	−2.240	When I read I have the impression that I have missed something important and must go back and reread the passage at least two or three times.
	2: Black	1.08	1.24	0.544	0.141			
43	1: White	0.44	0.84	1.784	0.069	0.080	0.472	I worry about remembering completely unimportant things and make an effort not to forget them.
	2: Black	0.51	0.97	1.864	0.155			
Factor 3: Mental Control/Doubt								
11	1: White	1.35	1.23	0.168	0.064	0.061	0.368	When doubts and worries come to my mind, I cannot rest until I have talked them over with a reassuring person.
	2: Black	1.40	1.33	0.229	0.153			
12	1: White	0.92	1.10	0.785	0.063	0.013	0.080	When I talk, I tend to repeat the same things and the same sentences several times.
	2: Black	0.94	1.15	0.798	0.149			
13	1: White	0.64	0.95	1.261	0.067	−0.145	−0.878	I tend to ask people to repeat the same things to me several times consecutively, even though I [understood what they said].
	2: Black	0.79	1.12	1.116	0.151			
18	1: White	0.53	0.87	1.569	0.077	0.222	1.139	I have to do things several times before I think they are properly done.
	2: Black	0.51	0.95	1.791	0.179			
26	1: White	1.07	1.17	0.588	0.060	0.298	1.877	I find it difficult to make decisions, even about unimportant matters.
	2: Black	0.86	1.10	0.886	0.147			
27	1: White	0.93	1.06	0.712	0.050	−0.062	−0.480	Sometimes I am not sure I have done things which in fact I know I have done.
	2: Black	1.02	1.10	0.650	0.119			

(continued)

TABLE 4 (continued)

Item	Group	Raw Score M	SD	b (location)	SE	DIF: $b_2 - b_1$	Critical Ratio	Question
28	1: White	1.04	1.20	0.649	0.057	−0.025	−0.172	I have the impression that I will never be able to explain things clearly, especially . . . about important matters that involve me.
	2: Black	1.08	1.28	0.624	0.134			
29	1: White	0.89	1.07	0.796	0.048	−0.059	−0.481	After doing something carefully, I still have the impression I have either done it badly or not finished it.
	2: Black	0.90	1.14	0.737	0.113			
30[a]	1: White	0.46	0.89	1.823	0.082	−0.478	−2.672	I am sometimes late because I keep on doing certain things more often than necessary.
	2: Black	0.73	1.20	1.345	0.159			
31	1: White	0.83	1.11	0.912	0.049	0.067	0.521	I invent doubts and problems about most of the things I do.
	2: Black	0.83	1.18	0.979	0.119			
32	1: White	1.04	1.16	0.601	0.055	0.190	1.312	When I start thinking of certain things, I become obsessed with them.
	2: Black	0.95	1.26	0.791	0.134			
33	1: White	1.01	1.22	0.707	0.054	−0.264	−1.952	Unpleasant thoughts come into my mind against my will and I cannot get rid of them.
	2: Black	1.17	1.38	0.443	0.124			
35	1: White	0.78	1.07	1.078	0.066	0.021	0.126	My brain constantly goes its own way, and I find it difficult to attend to what is happening round me.
	2: Black	0.87	1.25	1.099	0.153			
36	1: White	0.62	0.97	1.357	0.063	0.208	1.257	I imagine catastrophic consequences as a result of absent-mindedness or minor errors which I make.
	2: Black	0.61	1.05	1.565	0.153			
37	1: White	0.75	1.09	1.149	0.064	0.224	1.336	I think or worry at length about having hurt someone without knowing it.
	2: Black	0.67	1.10	1.373	0.155			
39	1: White	0.48	0.93	1.885	0.097	−0.147	−0.664	I sometimes worry at length for no reason that I have hurt myself or have some disease.
	2: Black	0.64	1.17	1.738	0.199			
42[a]	1: White	0.87	1.08	0.850	0.061	−0.355	−2.353	When I read I have the impression that I have missed something important and must go back and reread the passage . . . 2-3x.
	2: Black	1.08	1.24	0.495	0.138			
43	1: White	0.44	0.84	1.713	0.067	0.111	0.657	I worry about remembering completely unimportant things and make an effort not to forget them.
	2: Black	0.51	0.97	1.824	0.155			
44	1: White	0.74	1.04	1.126	0.064	−0.208	−1.312	When a thought or doubt comes into my mind, I have to examine it from all points of view and cannot stop until I have done so.
	2: Black	0.94	1.24	0.918	0.145			
45	1: White	1.02	1.17	0.636	0.059	−0.050	−0.333	In certain situations, I am afraid of losing my self-control and doing embarrassing things.
	2: Black	1.13	1.40	0.586	0.138			
52	1: White	0.55	1.00	1.697	0.087	0.135	0.627	I sometimes feel something inside me which makes me do things which are really senseless and which I do not want to do.
	2: Black	0.53	1.03	1.832	0.197			
59	1: White	0.60	0.95	1.498	0.086	0.245	1.125	When I hear about a suicide or a crime, I am upset for a long time and find it difficult to stop thinking about it.
	2: Black	0.57	1.05	1.743	0.200			
Factor 4: Fear of Impulses								
34[a]	1: White	0.71	1.17	1.236	0.069	−0.466	−2.870	Obscene or dirty words come into my mind and I cannot get rid of them.
	2: Black	0.93	1.37	0.770	0.147			
35	1: White	0.78	1.07	0.922	0.055	−0.186	−1.353	My brain constantly goes its own way, and I find it difficult to attend to what is happening round me.
	2: Black	0.87	1.25	0.736	0.126			
38	1: White	0.31	0.75	2.021	0.092	0.119	0.534	When I hear about a disaster, I think it is somehow my fault.
	2: Black	0.25	0.68	2.140	0.203			

(continued)

TABLE 4 (continued)

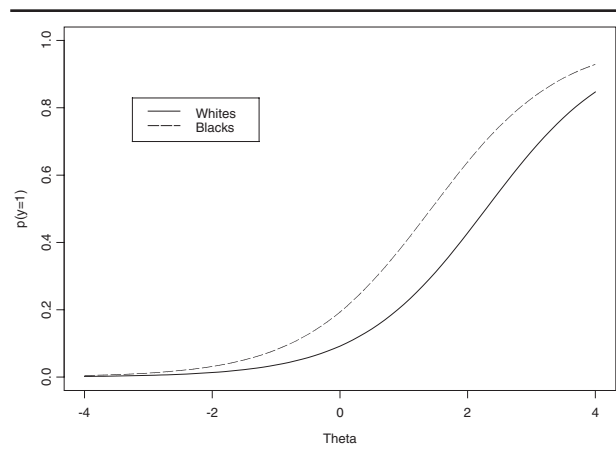| Item	Group	Raw Score M	SD	b (location)	SE	DIF: $b_2 - b_1$	Critical Ratio	Question
46	1: White	0.51	0.93	1.452	0.067	−0.041	−0.245	When I look down from a bridge or a very high window, I feel an impulse to throw myself into space.
	2: Black	0.55	1.08	1.411	0.153			
47	1: White	0.27	0.73	2.143	0.100	0.050	0.217	When I see a train approaching, I sometimes think I could throw myself under its wheels.
	2: Black	0.25	0.72	2.193	0.207			
48[a]	1: White	0.29	0.72	2.000	0.094	0.788	2.775	At certain moments, I am tempted to tear off my clothes in public.
	2: Black	0.15	0.66	2.788	0.268			
49	1: White	0.56	1.00	1.421	0.072	0.144	0.768	While driving, I sometimes feel an impulse to drive the car into someone or something.
	2: Black	0.48	0.96	1.565	0.173			
50	1: White	0.28	0.80	2.144	0.099	0.070	0.302	Seeing weapons excites me and makes me think violent thoughts.
	2: Black	0.28	0.91	2.214	0.120			
52	1: White	0.55	1.00	1.346	0.059	−0.111	−0.758	I sometimes feel something inside me which makes me do things which are really senseless and which I do not want to do.
	2: Black	0.53	1.03	1.235	0.134			
53	1: White	0.53	0.93	1.282	0.051	0.028	0.210	I sometimes feel the need to break or damage things for no reason.
	2: Black	0.51	1.01	1.310	0.123			
54	1: White	0.26	0.76	2.184	0.102	−0.113	−0.507	I sometimes have an impulse to steal other people's belongings, even if they are of no use to me.
	2: Black	0.24	0.69	2.071	0.198			
55	1: White	0.25	0.70	2.148	0.099	−0.103	−−0.473	I am sometimes almost irresistibly tempted to steal something from the supermarket.
	2: Black	0.30	0.82	2.045	0.194			
56	1: White	0.16	0.61	2.830	0.170	−0.217	−0.673	I sometimes have an impulse to hurt defenseless children or animals.
	2: Black	0.16	0.57	2.613	0.274			
57	1: White	0.43	0.86	1.661	0.080	0.038	0.195	I feel I have to make special gestures or walk in a certain way.
	2: Black	0.42	0.84	1.699	0.178			

a. These items showed significant DIF at $p < .05$

example, in a behavioral treatment study, Hatch, Friedman, and Paradis (1996) documented greater reluctance among Black clients to disclose OCD symptoms out fear of being labeled "crazy." Therefore, self-presentation bias may be a factor in the underendorsement of certain pathological traits by Blacks (Whaley, 2001). Overendorsement of certain items by Blacks could also be a means of positive self-presentation to counter negative stereotypes (Devine, 1989; Lepore & Brown, 1997). Our laboratory is currently trying to elucidate the mechanisms of racial DIF in anxiety items.

Limitations of This Study

There are several limitations to this study that warrant caution in interpreting the results. Although Internet data may improve response veracity due to enhanced anonymity, it may be more difficult to ensure that respondents are answering certain demographic questions truthfully. Data cleaning was essential, although for our sample it was not difficult to identify duplicates (matching demographic

FIGURE 1
Item Characteristic Curves for Black and White Participants on Item 10

variables and IP addresses) and surveys taken by nonserious responders. Like traditional student samples,

Internet samples are not representative, which may limit the generalizability of the findings. Participants are self-selected, not chosen at random, although recent research seems to indicate that self-selected Internet participants are at least as likely to give honest responses and produce valid data as traditional samples (Gosling, Vazire, Srivastava, & John, 2004). Overall, we feel these issues have not subtracted from the quality of the data, given the consistency in PI mean scores, standard deviations, and factor structure as compared to the findings of other studies using student and community samples. See Nosek, Banaji and Greenwald (2002) for a more in-depth discussion of issues involving research on the Internet, an increasingly important modality of scientific inquiry.

It should be noted that in this study, the middle response category of the PI was reworded for clarity; it has not been previously documented as to how this may affect the findings. Future studies should consider adopting the same or similarly clarified wording, because the original wording could be confusing to some English-speaking participants.

Although the causes of DIF for many items may result from cultural difference or self-presentation bias, more research is needed to confirm these hypotheses. One way to understand these differences would be to explore the meaning of items with those taking the test on an individual basis. It would also be useful to design measures that combine items describing classical OCD symptoms with items designed to measure cultural attitudes about cleanliness, normal worry, and control of behavior, so the relations between these dimensions and pathological traits can be better understood.

Summary and Future Directions

There are significant differences in the way Blacks and Whites respond to OCD questionnaires, especially in the areas of contamination fears and checking. These differences go beyond simple group differences in the level of the trait; evidence suggests that existing OCD items do not measure the same traits in Blacks and Whites. Such instruments may also exhibit bias among other groups, such as Hispanics. It seems clear that measures of OCD need to be evaluated with ethnic considerations in mind. The current findings are compelling and indicate an urgent need for more work in cross-cultural validation of assessment tools.

NOTES

1. Included in this number are 18 underrepresented minorities (Pacific Islanders, Native American) whose data were not used for individual group comparison.

2. The unadjusted group difference in mean b values for Blacks (b_2) and Whites (b_1) on this factor is $b_2 - b_1 = 0.891 - 1.331 = -0.440$.

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: American Psychiatric Press.

Azocar, F., Arean, P., Miranda, J., & Munoz, R. F. (2001). Differential item functioning in a Spanish translation of the Beck Depression Inventory. *Journal of Clinical Psychology*, *57*(3), 355-365.

Burns, G. L., Keortge, S. G., Formea, G. M., & Sternberger, L. G. (1996). Revision of the Padua Inventory of obsessive compulsive disorder symptoms: Distinctions between worry, obsessions, and compulsions, *Behaviour Research and Therapy*, *34*(2), 163-173.

Burns, G. L., Formea, G. M., Keortge, S. G., & Sternberger, L. G. (1995). The utilization of non-patient samples in the study of obsessive-compulsive disorder. *Behaviour Research and Therapy*, *33*(2), 133-144.

Cooke, D. J., Kosson, D. S., & Michie, C. (2001). Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist–Revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment*, *13*(4), 531-542.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5-18.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Hillsdale, NJ: Lawrence Erlbaum.

Fabrega, H., Mezzich, J., & Ulrich, R. F. (1988). Black-White differences in psychopathology in an urban psychiatric population. *Comprehensive Psychiatry*, *29*(3), 285-297.

Feske, U., & Chambless, D. L. (2000). A review of assessment measures for obsessive-compulsive disorder. In W. K. Goodman, M. V. Rudorfer, & J. D. Maser (Eds.), *Obsessive-compulsive disorder: Contemporary issues in treatment. Personality and clinical psychology series* (pp. 157-182). Hillsdale, NJ: Lawrence Erlbaum.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, *59*(2), 93-104.

Hafner, R. (1988). Obsessive-compulsive disorder: A questionnaire survey of a self-help group. *International Journal of Social Psychiatry*, *34*(4), 310-315.

Hatch, M. L., Friedman, S., & Paradis, C. M. (1996). Behavioral treatment of obsessive-compulsive disorder in African Americans. *Cognitive & Behavioral Practice*, *3*(2), 303-315.

Hodgson, R. J., & Rachman, S. (1977). Obsessional-compulsive complaints. *Behaviour Research and Therapy*, *15*, 389-395.

Karno, M., Golding, J. M., Sorenson, S. B., & Burnam M. A. (1988). The epidemiology of obsessive compulsive disorder in five US communities. *Archives of General Psychiatry*, *45*, 1094-1099.

Kyrios, M., Bhar, S., & Wade, D. (1996). The assessment of obsessive-compulsive phenomenon: psychometric and normative data on the Padua Inventory from an Australian non-clinical student sample. *Behaviour Research and Therapy*, *34*(1), 85-95.

Kyrios, M., & Iob, M. (1998). Automatic and strategic processing in obsessive-compulsive disorder: Attentional bias, cognitive avoidance or more complex phenomena? *Journal of Anxiety Disorders*, *12*(4), 271-292.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*, 275-287.

Muraki, E. (1990). Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement*, *14*, 59-71.

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, *17*, 351-363.

Muraki, E., & Bock, R. D. (1993). *PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks*. Chicago: Scientific Software International.

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide, version 2*. Los Angeles: Author.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). eResearch: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues*, *58*(1), 161-176.

Orlando, M., & Marshall, G. N. (2002). Differential item functioning in a Spanish translation of the PTSD Checklist: Detection and evaluation of impact. *Psychological Assessment*, *14*(1), 50-59.

Ritsher, J. B., Stuening, E. L., Hellman, F., & Guardino, M. (2002). Internal validity of an anxiety disorder screening instrument across five ethnic groups. *Psychiatry Research*, *111*, 199-213.

Sanavio, E. (1988). Obsessions and compulsions: The Padua Inventory. *Behaviour Research and Therapy*, *26*(2), 169-177.

Sternberger, L. G., & Burns, G. L. (1991). Obsessive compulsive disorder: Symptoms and diagnosis in a college sample. *Behavior Therapy*, *22*(4), 569-576.

Thomas, J., Turkheimer, E., & Oltmanns, T. F. (2000). Psychometric analysis of racial differences on the Maudsley Obsessional Compulsive Inventory. *Assessment*, *7*(3), 247-258.

Valleni-Basile, L. A., Garrison, C. Z., Waller, J. L., Addy, C. L., McKeown, R. E., Jackson, K. L., et al. (1996). Incidence of obsessive-compulsive disorder in a community sample of young adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(7), 898-906.

Van Oppen, P. (1992). Obsessions and compulsions: Dimensional structure, reliability, convergent and divergent validity of the Padua Inventory. *Behaviour Research & Therapy*, *30*(6), 631-637.

Whaley, A. L. (2001). Cultural mistrust of White mental health clinicians among African Americans with severe mental illness. *American Journal of Orthopsychiatry*, *71*(2), 252-256.