Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

3-2008

A Cross-Nations, Cross-Cultures, and Cross-Conditions Analysis on the Equivalence of the Balanced Inventory of Desirable Responding (BIDR)

Andrew LI
West Texas A&M University

Jochen REB
Singapore Management University, jochenreb@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Organizational Behavior and Theory Commons

Citation

LI, Andrew and REB, Jochen. A Cross-Nations, Cross-Cultures, and Cross-Conditions Analysis on the Equivalence of the Balanced Inventory of Desirable Responding (BIDR). (2008). *Journal of Cross-Cultural Psychology*. 40, (2), 214-233.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/1011

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

A Cross-Nations, Cross-Cultures, and Cross-Conditions Analysis on the Equivalence of the Balanced Inventory of Desirable Responding (BIDR)

Andrew Li

West Texas A&M University

Jochen Reb

Singapore Management University

Please address correspondence to:

Andrew Li

WTAMU Box 60809

Canyon TX 79016

Email: ali@WTAMU.edu

Abstract

This study examined measurement equivalence of the Balanced Inventory of Desirable Responding (BIDR) across two nations: the United States and Singapore; two cultural values: Horizontal Individualism (HI) and Horizontal Collectivism (HC); and two motivational conditions: standard and faking. One sample of undergraduate students from each country ($N_{Singapore} = 158$; $N_{United\ States} = 166$) participated in this study. A within-subject experimental design was used in this study. Specifically, at Time I, participants were simply asked to respond to the BIDR and the INDCOL (standard condition). At Time II, they were instructed to engage in social desirability (faking condition). Multi-group Confirmatory Factor Analyses were used to evaluate the equivalence of the BIDR. We found support for the equivalence of the BIDR across the two cultural values. However, there was weaker support for the equivalence of the BIDR across the two countries and the two motivational conditions. The implications of these findings are discussed.

Keywords: BIDR, Measurement Equivalence, Cross-Cultural Research, Social Desirability

In recent years, personality inventories have enjoyed a resurgence as a selection tool. However, one persistent criticism of the use of personality scales in the selection context is that they may be vulnerable to social desirability (Edwards, 1957). Social desirability, or response distortion, is defined as “the tendency to endorse items in response to social or normative pressures instead of providing veridical self-reports” (p. 122, Ellingson, Smith, & Sackett, 2001). Social desirability may influence the ranking of job applicants such that those who distort their responses end up ranking higher than those who respond to the scales honestly (Mueller-Hanson, Heggestad, & Thornton, 2003). As such, social desirability may unfairly disadvantage those applicants who respond honestly to personality inventories (Li & Bagger, 2006).

These concerns have led researchers to develop validity scales to identify applicants suspected of providing socially desirable responses (Li & Bagger, 2006). One of the most widely used validity scales is the Balanced Inventory of Desirable Responding (BIDR, Paulhus, 1991, cf. Li & Bagger, 2007). Researchers have found that job applicants scored higher on the BIDR than job incumbents, attesting to the use of social desirability in the selection process (e.g., Rosse, Stecher, Miller, & Levin, 1998). However, one assumption behind this comparison is that job applicants’ BIDR scores are compatible with incumbents’ BIDR scores (Vandenberg & Lance, 2000). However, since job applicants and job incumbents are differently motivated to engage in social desirability, it is not clear whether the BIDR completed by these two groups is equivalent. Without evidence of equivalence, comparisons of the BIDR scores may be potentially erroneous or misleading. Additionally, besides being used widely in North America, the BIDR is increasingly being employed in other countries (e.g., Australia, Israel, China, Singapore, Fox & Schwartz, 2002; Geiger & O’Connell, 2000; Lay et al., 1998; Stober, 2001). Unfortunately, none of these studies examined whether the BIDR is cross-culturally or crossnationally invariant. As noted by cross-cultural scholars, the lack of evidence for measurement invariance may lead to concerns about whether the BIDR captures the same construct when used in another culture, and may raise doubts about the results of cross-cultural comparisons (Vandenberg & Lance, 2000).

To address these concerns, we conducted a cross-nations, cross-cultures, and cross-conditions analysis of the equivalence of the BIDR. For the cross-nations analysis, we examined the measurement equivalence of the BIDR across two countries: the United States and Singapore. Given research showing that nation is often not a good surrogate for culture (Lam, Schaubroeck, & Aryee, 2002), we also examined the equivalence of the BIDR across two cultural dimensions: horizontal individualism and horizontal collectivism. Finally, to investigate the transportability of the BIDR as a function of one's motivation to engage in social desirability, we examined the equivalence across two different motivational conditions: standard and faking.

The BIDR

The popularity of the BIDR can be in part attributed to the fact that it is one of the few validity scales that have subscales separately tapping two different forms of social desirability: self-deception enhancement and impression management (Paulhus, 1984). *Self-deception enhancement* (SDE) refers to an unintentional tendency to overestimate one's strengths, manifested in honestly believed, but inaccurate, self descriptions. *Impression management* (IM), in contrast, refers to a deliberate attempt to deceive others by painting one's public image in an overly positive light. Thus, self-deception is manifested as a tendency to deny universally experienced thoughts or behaviors (e.g., of a sexual or aggressive nature) in order to protect one's self-image, whereas impression management is manifested as an inclination to claim desirable (but rather infrequent) behaviors with the purpose of projecting a favorable public image.

Measurement Equivalence

Measurement invariance/equivalence is defined as "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attributes" (Horn & McArdle, 1992, p. 117). Absent evidence of measurement invariance, it is difficult to assess whether the measure captures the same underlying constructs across different cultures/nations/conditions. As such, comparisons on means, correlations, variance, or covariance may yield results that are difficult to interpret or even outright misleading. As Vandenberg and Lance (2000) noted, "if not tested, violations of measurement equivalence assumptions are as threatening to substantive interpretations as is an inability to demonstrate reliability and validity" (p. 6).

Cross-Cultures Equivalence

Cultures differ in a number of ways. One way to characterize these differences is through an individual's relationship with his/her group, termed as individualism/collectivism (I/C). According to Hofstede (2001), in countries that are high on individualism, people tend to see themselves as being independent from each other. The cultural mission is to express one's uniqueness and to achieve personal goals. In contrast, in countries that are high on collectivism, people tend to consider themselves as a fabric of a social network. Therefore, they are motivated to achieve group goals and to maintain social harmony.

In recent years, researchers argued that I/C also has a second dimension: horizontal versus vertical (Singles, Triandis, Bhawuk, & Gelfand, 1995; Triandis & Gelfand, 1998). The horizontal dimension focuses on equal status, whereas the vertical dimension places more emphases on hierarchy. These two dimensions cross to produce four cultural orientations: Horizontal Individualism (HI), Horizontal Collectivism (HC), Vertical Individualism (VI), and Vertical Collectivism (VC). Both HC and VC emphasize the interdependence of social relationships. However, horizontals believe in equality in the society; as such, although they are willing to exert effort to build interpersonal relationships and achieve common goals, they are less willing to do so through submission to authorities. In contrast, verticals believe in the hierarchy of social arrangements. Therefore, they are willing to accept orders from an authority for the sake of achieving group goals. Similarly, both HI and VI share the emphasis on independence and personal goals. However, horizontals are more interested in expressing their individuality while verticals are more interested in acquiring social status and personal success through competitions.

Several researchers have argued that people in collectivistic cultures are more likely to engage in impression management. As Triandis and Suh (2002, p. 144) pointed out, “lying is a more acceptable behavior in collectivist than in individualist cultures.” They reasoned that since group harmony is important in collectivistic cultures, people may lie to save each other’s face and avoid direct confrontations. In other words, when lying is in the service of maintaining group harmony, it is more likely to be condoned or even encouraged. Consistent with this argument, Triandis and colleagues (2001) found that the tendency to deceive was greater among collectivist countries than individualistic countries. In contrast, people in individualistic cultures are more likely to inflate their score on self-deception. According to Markus and Kitayama (1991), people from individualistic cultures tend to use their internal attributes for self definition. Therefore, they are more likely to engage in self-deception to maintain a positive self view.

Lalwani, Shavitt, and Johnson (2006) argued that the horizontal/vertical dimension may further refine the relationship between I/C and social desirability. According to Lalwani et al. (2006), many of the items in the SDE measure tap the tendency to inflate one's scores on self-determination and uniqueness, rather than personal success. Since individuals high on HI value independence and self-reliance, they are more likely to inflate their scores on SDE to project an image of being "reliant, skillful, and independent" (p. 174). In contrast, people who are high on VI are more focused on individual achievement and personal success and less on uniqueness and self-determination. Thus they are less likely to engage in SDE.

Lalwani and colleagues (2006) also argued that many items in the IM measure focus on maintaining interpersonal relationships. Those scoring high on HC emphasize the importance of social relationships. Therefore, they may be more motivated to inflate their scores on the IM to appear cooperative, sociable, benevolent, and loyal. In contrast, those who score high on VC may be more concerned with conformity to authority; as such, they may be less likely to inflate their scores on the IM. Lalwani and colleagues (2006) found that while those scoring high on HI also tended to score high on SDE, no such relation was found for VI. Further, consistent with their hypothesis, those who scored high on HC also tended to score high on IM, but no such relation was found for VC.

These theoretical arguments and empirical findings suggest that people scoring high versus low on HI and HC may respond differently to IM and SDE. As such, we examined the equivalence of the BIDR across groups formed based on their scores on HI and HC.

Cross-Nations Equivalence

Although the BIDR has been developed and validated in North America, in recent years it has been used in research in other countries (e.g., Fox & Schwartz, 2002; Lay et al., 1998; Stober, 2001). However, to our knowledge, there has been no systematic effort to assess measurement equivalence of the BIDR in different countries. Thus the second goal of this study was to examine the cross-nations measurement equivalence of the BIDR, involving participants from the United States and Singapore.

The United States and Singapore were chosen for several reasons. First, these two countries are different from each other culturally. In the GLOBE study (House, Hanges, Javidan, Dorfman, & Gupta, 2004), these two countries were classified into different bands based on seven out of the nine societal values (countries that were classified into separate bands are considered truly different from each other on that particular cultural value). Specifically, the United States and Singapore belonged in different bands based on the cultural values of performance orientation, future orientation, gender egalitarianism, assertiveness, institutional collectivism, power distance, and human orientation. These cultural differences may elicit differences in the way people interpret the connotations of the items in the measure (Robert, Lee, & Chan, 2006). Second, research has suggested that people from different countries may calibrate rating continuums differently. For example, past research has suggested that compared with Americans, East Asians are less willing to endorse the extreme points of a rating continuum (c.f., Cheung & Rensvold, 2000). Such differences may create threat to scale equivalence.

Third, although English is the official language of both countries, given that the BIDR was originally designed for North American subjects, some of the idiomatic expressions in the measure may not be accessible to people from other countries. For example, the BIDR includes an item “I sometimes lose out on things because I can’t make up my mind soon enough.” The phrase “lose out on things” may not be familiar to Singaporeans. Given this concern, Ryan, Chan, Ployhart, and Slade (1999) argued that “a similar language across versions does not guarantee measurement equivalence.” Thus, choosing these two English-speaking countries may allow us to evaluate the within-language equivalence of the BIDR.

Fourth, the comparison between these two countries also allows us to evaluate whether items in the scale may or may not be equally relevant to people from different countries (Byrne & Watkins, 2003). For example, the BIDR includes items such as "I sometimes drive faster than the speed limit" and "I am not a safe driver when I exceed the speed limit." These items may be more appropriate in the United States (which is a four-wheel nation), and less so in other countries, such as Singapore, where most people rely on public transportations.

It is important to note the difference between the cross-nations comparison discussed in this section and the cross-cultures comparison discussed in the previous section. Most of the studies on measurement equivalence equate national difference with cultural difference, such that detection of measurement non-equivalence across nations is automatically attributed to cultural difference (e.g., Ghorpade, Hattrup, & Lackritz, 1999; Liu et al., 2004). However, recent research has started to question these assumptions. For example, Ghopade and colleagues (1999) argued the importance of "conceptualizing culture as a construct that varies among groups within a country" (p. 678). Clearly, if one suspects that it is cultural difference, rather than national difference per se, that leads to measurement equivalence, then it would be useful to examine equivalence between groups that are formed based on differences in the relevant cultural dimensions. However, as we discussed in this section, culture is hardly the only mechanism that may lead to measurement non-equivalence. Thus, it is important to examine the equivalence of the BIDR across these two different nations.

Cross-Conditions Equivalence

One often cited piece of evidence for the presence of social desirability in personality tests is that job applicants score higher on validity scales such as the BIDR than do incumbents

(Rosse et al., 1998). One underlying assumption of such comparisons is that the BIDR means the same thing for both groups. However, this assumption has never been tested empirically. There are several reasons to expect that motivations to engage in social desirability may influence measurement equivalence. First, researchers have shown that social desirability may change the structure of personality measures. Specifically, the correlations among underlying factors tend to increase with job applicant samples, thus changing the factor structure of these measures. For example, Ellingson, Sackett, and Hough (1999) found that instructing people to respond in a socially desirable manner resulted in a uni-dimensional structure of a personality measure that was multi-dimensional in a less-motivated sample. While direct evidence is scant, researchers have found that the correlation between the IM and the SDE increased when the motivational level of the sample was high (for example, job applicant sample), thus raising questions about the transportability of the BIDR structure across samples with varying motivations to engage in social desirability (Barrick & Mount, 1996).

Second, people with different motivations to engage in social desirability may respond differently to the same items as a function of their favorability. For example, an item like "when my emotions are aroused, it biases my thinking" may only elicit strong desire to inflate scores among people who are motivated to fake. Therefore, people with equivalent standing on an item like this may respond differently based on their motivation to engage in social desirability.

As such, the third goal of the present study was to examine measurement equivalence of the BIDR under two different motivational contexts: a standard condition, in which no particular information on how to respond was given, and a faking condition, in which participants were instructed to provide socially desirable responses.

Method

Sample

The total sample included 336 undergraduate students, with 172 of them from a university in the southwestern part of the United States and 164 from a university in Singapore. Elimination of missing data reduced the number of participants to 324 (166 from the United States and 158 from Singapore). All participants from both countries indicated that English was their first language. Demographic information for these two samples is presented in Table 1.

Design and Procedure

Participation of this study was completely voluntary. Participants earned extra credit as well as a chance to earn cash rewards (to be discussed below). The experiment was conducted outside of class time. Participants arrived in a large lecture hall, where they were individually seated. We used a within-subject design with each participant completing the BIDR twice: once in the standard condition and once in the faking condition. In the first administration, the *standard* condition, participants were simply instructed to respond to the BIDR and the INDCOL by using the rating scale provided. Two weeks later, participants returned for the second administration, the *faking* condition, in which they were instructed that they were taking this test as part of the job application process. They should do their best to enhance their chance of being hired. The instructions given to the participants at Time 2 were as follows:

> *Imagine yourself taking this test when you apply for a job that you really like. Respond to these items in a way that will increase your chance of being selected.*
>
> *To encourage good performance in this test, those who score at the top 10% among all test-takers will be entered in a drawing for 4 $25 cash prizes.*

The standard condition was run before the faking condition for all participants (rather than counterbalancing the presentation order) because past research has shown that a faking condition may be more likely to undermine the subsequent honest responses than the other way around (Holden, 1997). Also, an interval of two weeks was chosen because scores in the second administration may be contaminated by the scores in the first administration, if the two administrations are too temporally adjacent (Ellingson et al., 1999). Having an interval of two weeks mitigated the risk of a repeated measures effect.

Measures

BIDR. Participants completed the 6th version of the BIDR in both administrations (Paulhus, 1991). Each of the two subscales, IM and SDE, includes 20 items. A sample item of IM is “I have received too much change from a sales person without telling him or her.” A sample item of SDE is “I never regret my decision.” Respondents were instructed to rate on a 1-7 continuum the degree to which each statement applied to them (1 anchors “strongly disagree” and 7 anchors “strongly agree”). Half of the questions in the BIDR are keyed in the negative direction (Paulhus, 1991). The scores of these items were reverse-coded.

HI and *HC*. Participants in the standard condition also completed the INDCOL (Singelis et al., 1995). The INDCOL is a popular scale used to measure individualism/collectivism (I/C). One notable difference between the INDCOL and other I/C scales is that it also includes the horizontal/vertical dimension. HI and HC each have eight items. A sample item of the HI is “I like my privacy.” A sample item of the HC is “I feel good when I cooperate with others.”

Results

Tables 2 (U.S.) and 3 (Singapore) show the means, standard deviations, reliability coefficients, and inter-correlations of all measured variables. It is worth noting that higher scores on the IM and SDE were reported in the faking condition than the standard condition for both samples. The mean of IM scores for the American sample was 3.6 (men: 3.5; women: 3.7), which was lower than the mean score of 4.8 (men: 4.5; women: 5.2) reported in a different study (Bernardi, 2006).

The test of measurement equivalence includes several critical steps (Cheung & Rensvold, 2002; Little, 1997; Vandenberg & Lance, 2000). First, it is important to establish *configural* invariance. In other words, items should load on the same factors. According to Vandenberg and Lance (2000), configural invariance is the pre-requisite for subsequent steps of measurement equivalence testing. This also represents the baseline model against which all other more restricted models are compared. Second, it is also important to examine whether factor loadings (slopes) for the same item are equivalent (*metric* invariance). Third, an additional step is the evaluation of the equivalence of intercepts (*scalar* invariance). In the fourth step, the equivalence of measurement error is evaluated (*error variance* invariance).

We used the multi-group structural equation modeling procedure recommended by Byrne (2004) to test measurement equivalence of the BIDR. In essence, we compared the baseline model with a series of progressively restrictive models. Model fit was evaluated by the following indexes: the relative Chi-square index (χ^2/df), the Comparative Fit Index (CFI), the Normed Fit Index (NFI), the Root Mean Squared Error of Approximation (RMSEA), and Standard Root Mean Square Residual (SRMR). According to Carmines and McIver (1981), a relative Chi-square index within the range of 2:1 or 3:1 can be considered a good fit (the larger the value, the worse fit). CFI and NFI are fit indexes that compare the fit of a model with a baseline model in which there is no correlation or covariance among variables (Bentler, 1990). CFI and NFI estimates of equal to or greater than .90 are indicative of good fit (the lower the value, the poorer the fit). RMSEA provides inference about the discrepancy between a measured model and a saturated model. A value of .08 or less for RMSEA is considered acceptable fit, a value between

.08 and .10 indicates moderately acceptable fit, and a value of .10 or above is considered a poor fit (the higher the value, the worse fit). SRMR provides information about the standardized difference between observed and predicted variance and covariance. A value of .08 or less for SRMR indicates good fit (the higher the value, the worse fit).

Model invariance is determined by CFI difference as well as Chi-square difference ($\Delta\chi^2$). Researchers argue that the use of $\Delta\chi^2$ alone to determine the significance of model change is problematic because Chi-square values are overly sensitive to misfit. Cheung and Rensvold (2002) recommended that CFI difference of .01 or less can be considered model invariance. Therefore, we consider the difference between two models as being significant when both conditions are met ($\Delta\chi^2$ is significant and CFI difference is greater than .01).

Cross-Nations Analysis

In the first step, we examined the equivalence of the BIDR across two countries: the United States and Singapore. Results of these analyses are presented in Table 4. Before we conducted these analyses, we also examined the difference of reliability coefficients across these two nations. The differences were not statistically significant: Standard (IM: $F(165, 157) = 1.17$, ns; SDE: $F(165, 157) = .70$, ns); Faking (IM: $F(165, 157) = 1$, ns; SDE: $F(165, 157) = .70$, ns).

Models 1 and 2. The first two models examined whether the hypothesized two-factor structure (IM and SDE) adequately fit the data in each sample and for each response condition. We specified each item to load on the theoretically expected factor only (no cross-loading). One item was randomly chosen from each factor and the loading of this item was fixed to one; the loadings of other items were allowed to be freely estimated (cf. Byrne, 2004). Both models in the faking condition showed a good fit. In the standard condition, most of the indexes were within the acceptable range, with the exception of SRMR for both models.

Model 3. In the next step, we constrained the model so that the items loaded on the same factor across samples (configural invariance). The overall model fit was adequate in both response conditions (the SRMR index was slightly higher than the cutoff for the standard condition).

Model 4. Next, in Model 4, we constrained for equal *factor loading (slope)* for each item, and compared this constrained model with Model 3. Although the $\Delta\chi^2$ between the constrained Model 4 and the baseline Model 3 was significant (standard: $\Delta\chi^2 = 81.67$, df = 38, p < .05; faking: $\Delta\chi^2 = 60.83$, df = 38, p < .05), the CFI difference was within .01 for both response conditions. Thus, we concluded that constraining slopes to be equal did not lead to a significant decrease in model fit.

Model 5. In Model 5, we constrained slopes and *intercepts* to be equal across the American and Singaporean samples. In the standard condition, the constraints led to a CFI change of .011 (between Model 5 and Model 3, Table 4), which was greater than the .01 threshold recommended by Cheung and Rensvold (2002). Additionally, the $\Delta\chi^2$ was also significant ($\Delta\chi^2 = 360.75$, df = 76, p < .05). These results suggest that constraining the intercept to be equal across both groups led to measurement non-equivalence. In the faking condition, the CFI change from Model 3 to Model 5 was less than .01 ($\Delta\chi^2$ was significant, $\Delta\chi^2 = 359.04$, df = 76, p < .05), indicating that the measure showed scalar invariance across these two countries.

Model 6. In Model 6 we constrained slope, intercept, and *error variance* to be equal across both samples. For the standard condition, the CFI difference between Model 6 and Model 3 exceeded the .01 cutoff. Additionally, the $\Delta\chi^2$ was significant ($\Delta\chi^2 = 456.15$, df = 114, p < .05). Similarly, in the faking condition, the constraints on slope, intercept, and error variance led to a CFI difference of .013 between Model 3 and Model 6, which was greater than the .01 threshold.

The $\Delta\chi^2$ between the two models was also significant ($\Delta\chi^2$ = 588.93, df = 114, p < .05). These results suggest that these constraints led to non-equivalence of the BIDR across both conditions.

Cross-Cultures Analysis

We next examined the equivalence of the BIDR across two cultural values: HI and HC. In the first step of analysis, we evaluated the equivalence of HI and HC across the United States and Singapore. This step is necessary, as it allowed us to combine the data from these two countries and then to regroup them based on the HI and HC scores. Results of the analyses, presented in Table 5, show that all the fit indexes were satisfactory, indicating measurement equivalence of HI and HC across these two countries.

In the next step, we divided the entire sample (both the Americans and Singaporeans) into two groups based on the median of HI scores. Each group included 162 participants, with the high HI group including 49% Singaporeans, and the low HI group including 48% Singaporeans. The mean of HI for the high HI group was 5.86 (SD = .34), and the mean for the low HI groups was 4.95 (SD = .38). We then examined the measurement equivalence of the BIDR across these two groups with constraints placed on factor structure, slope, intercept, and error variance. Before we conducted these analyses, we examined the differences in reliability coefficients across the high and low HI groups. These comparisons were statistically insignificant: Standard (IM: F (161, 161) = 1.05, ns; SDE: F(161, 161) = 1.10, ns); Faking (IM: F (161, 161) = .90, ns; SDE: F(161, 161) = .95, ns).

Results of the cross-culture analyses are presented in Table 6. These results suggest the robustness of the BIDR across groups varying on HI scores. The CFI difference was always smaller than .01 across all the comparisons between the baseline model and the more restricted models. Additionally, these results were replicated across both the standard (Models 3 and 4: $\Delta\chi^2$

= 50.28, df = 38, *ns*; Models 3 and 5: $\Delta\chi^2$ = 182.46, df = 76, p < .05; Models 3 and 6: $\Delta\chi^2$ = 264.48, df = 114, p < .05) and the faking conditions (Models 3 and 4: $\Delta\chi^2$ = 42.35, df = 38, *ns*; Models 3 and 5: $\Delta\chi^2$ = 80.22, df = 76, *ns*; Models 3 and 6: $\Delta\chi^2$ = 149.22, df = 114, p < .05).

In the second set of analyses, we divided the entire sample into two groups based on HC scores. Each group included 162 participants, with the high HC group including 51% Singaporeans, and the low HC group including 47% Singaporeans. The mean of HC for the high HC group was 5.86 (SD = .35), and the mean for the low HC group was 4.80 (SD = .45). The equivalence of the BIDR was then evaluated across these two groups with a series of progressively restrictive models on factor structure, slope, intercept, and error variance. Again, before we conducted these analyses, we examined the differences in reliability coefficients. Most of these comparisons were statistically insignificant: Standard (IM: F (161, 161) = .97, ns; SDE: F(161, 161) = 1.44, P < .05); Faking (IM: F (161, 161) = .90, ns; SDE: F(161, 161) = 1.21, ns).

Results of these analyses, presented in Table 7, suggest the equivalence of the BIDR across these two groups. None of the constraints reduced the CFI appreciably beyond the baseline model (more than the .01 limit). These results were observed across both the standard (Models 3 and 4: $\Delta\chi^2$ = 43.67, df = 38, *ns*; Models 3 and 5: $\Delta\chi^2$ = 116.57, df = 76, p < .05; Models 3 and 6: $\Delta\chi^2$ = 186.8, df = 114, p < .05) and the faking conditions (Models 3 and 4: $\Delta\chi^2$ = 66.54, df = 38, p < .05; Models 3 and 5: $\Delta\chi^2$ = 109.17, df = 76, p < .05; Models 3 and 6: $\Delta\chi^2$ = 203.93, df = 114, p < .05). Taken together, these results support the equivalence of the BIDR across different values of HI and HC.

Cross-Conditions Analysis

We next examine the equivalence of the BIDR across two motivational conditions: standard and faking. We conducted these analyses separately for the American and Singaporean samples. We first examined the differences in reliability coefficients. All of the differences were statistically significant. US (IM: F (165, 165) = 1.50, p < .05; SDE: F(165, 165) = 1.46, p < .05); Singapore (IM: F (157, 157) = 1.91 p < .05; SDE: F(157, 157) = 1.37, p < .05). As can be seen in Table 8, for the American sample, the CFI difference between Model 4 (metric invariance) and Model 3 (the baseline model) was less than .01 ($\Delta\chi^2$ = 94.02, df = 38, p < .05). However, when constraints were placed on both factor loadings and intercepts, the difference between Model 5 and Model 3 exceeded the .01 threshold ($\Delta\chi^2$ = 540.95, df = 76, p < .05). Further, the CFI difference between Model 6 (with constraints on loadings, intercepts, and error variance) and Model 3 was also greater than .01 ($\Delta\chi^2$ = 1230.96, df = 114, p < .05). In all these comparisons, the $\Delta\chi^2$ was significant. We found the same results in the Singaporean sample. Specifically, although the CFI difference between Model 3 and Model 4 was less than .01 ($\Delta\chi^2$ = 140.92, df = 38, p < .05), the CFI differences between Model 5 and Model 3 ($\Delta\chi^2$ = 423.63, df = 76, p < .05), and Model 6 and Model 3 ($\Delta\chi^2$ = 610.44, df = 114, p < .05) were both greater than .01. Taken together, these results failed to support the equivalence of the BIDR across different motivational conditions.

Discussion

The purpose of this study was to examine the equivalence of the BIDR across two different cultural values (HI and HC), two different countries (Singapore and the United States), and two different motivational conditions (standard and faking). We found evidence for the equivalence of the BIDR across the two cultural values of HI and HC even when we constrained the factor loadings, intercepts, and error variance to be equal. Additionally, although we found support for metric invariance of the BIDR across the two nations, we failed to find evidence for scalar and error variance invariance. Similarly, while we found metric invariance of the BIDR across the two motivational conditions, no support was found for scalar and error variance invariance. In the sections below, we discuss the implications of these findings.

As a social desirability scale, the BIDR has been found useful in identifying intentional distortion to personality measures (Paulhus, 1991). These results prompted researchers to adopt the BIDR for their research in other cultural contexts. However, various researchers have cautioned that evidence of measurement equivalence has to be sought before a measure can be used in other cultures (e.g., Vandenberg & Lance, 2000). This caution is predicated on various threats to the generalizability of the BIDR across cultures (Robert et al., 2006). In the absence of evidence supportive of measurement equivalence, it is unclear whether the measure actually taps the intended construct, whether the underlying structure of the measure can hold up, or whether it is appropriate to conduct comparisons at the mean levels (Vandenberg & Lance, 2000).

Against this backdrop, we examined the equivalence of the BIDR used in two nations: the United States and Singapore. Although we found support for metric invariance, we found no evidence for scalar and error variance invariance. These findings raise important concerns for using the BIDR in cross-nations contexts. For example, several recent studies compared the means of the IM or the SDE across participants from multiple national backgrounds; but none of these studies tested for measurement equivalence prior to the mean comparisons (e.g., Lalwani et al., 2006). Without evidence for scalar invariance, such comparisons may yield results that are erroneous at best and misleading at worst (Steenkamp & Baumgartner, 1998).

One potential strategy to address this problem is to identify and exclude non-invariant items from mean score computations. This strategy is particularly useful when the non-invariant items only account for a minority of the entire scale and removing them may not damage the construct validity of the measure (Cheung & Rensvold, 1999, 2002). Chan (2000) suggested that as a precaution, when conducting mean comparisons, one may conduct parallel analyses, one with the non-invariant items and one without. This may allow researchers to evaluate whether removing the non-invariant items may have substantial impact on mean comparisons.

Past studies on measurement equivalence tended to equate nation with culture. Various researchers have criticized that this approach rests on the unwarranted assumption that groups from the respective nations are homogeneous in certain cultural dimensions. As Liu and colleagues (2004) pointed out, “there can be considerable cultural differences within a country, and different countries can be highly similar culturally” (p. 1080).

To address this limitation, we evaluated the equivalence of the BIDR across groups that were created based on the scores on two relevant cultural dimensions, Horizontal Individualism (HI) and Horizontal Collectivism (HC). Results of our analyses indicate that neither HI nor HC had much impact on the equivalence of the BIDR. These results raise an interesting question: what was the mechanism underlying the non-equivalence of the BIDR across the United States and Singapore, given that the cultural values of HI and HC did not seem to be the cause? One possibility is that some other cultural dimensions might have been responsible for the measurement non-equivalence observed in our cross-nations analysis. For example, Bernardi (2006) found that the cultural value of uncertainty avoidance was related to people’s tendency to engage in social desirability. Specifically, people from high uncertainty avoidance cultures were more likely to respond in a socially desirable manner than those from low uncertainty avoidance cultures. These findings suggest that future research should explore other cultural values, such as uncertainty avoidance, and examine their impact on the equivalence of the BIDR.

Another potential explanation for these findings may have to do with factors that reside within the national contexts. Nations are different from each other in various ways, and culture is merely one of them. For example, Tsui, Nifadkar, and Ou (2006) articulated a list of factors that distinguish different nations, including physical context (such as climate and geography), historical context (such as language and sovereignty), social context (such as religion and family structure), economic context (such as economic and business systems), and cultural context. Some of these contexts may lead to non-equivalence of the BIDR across nations. For example, the social and economic contexts of the United States may make the content of some items more concrete and relevant to Americans and less so to Singaporeans. Our earlier discussion of national differences in the use of cars as a means of transportation may provide a good example attesting to these effects. Linguistically, idiomatic expressions that are specific to one nation may also lead to misunderstanding and confusion when used in another nation. Given that the BIDR was developed in North America, it is possible that some of the items may contain expressions that are not readily accessible to Singaporeans. In sum, there are many factors that may lead to measurement non-equivalence across nations, and culture is merely one of them.

Finally, we found little support for the invariance of the BIDR across different motivational conditions. In other words, the BIDR completed under the standard condition was not equivalent to the BIDR under the faking condition. These results point to the risk of comparing BIDR scores when completed by people with varying motivations to engage in social desirability. As our results suggest, such comparisons may be potentially misleading, as the BIDR scores are not directly comparable without evidence of equivalence. These results, however, should be interpreted in light of the fact that, in the present study, social desirability was induced by explicit instructions to fake. As researchers have observed, such manipulations typically lead to score inflation across all the items (Ellingson et al., 2001). However, in other settings, such as the job application process, job applicants may be more subtle and respond in a more socially desirable manner to some items and less so to others. This may result in different patterns of social desirability across different settings (e.g., lab versus field). Future research should examine whether the BIDR may be equivalent across job applicants and incumbents.

Limitations and Future Research

The above findings need to be evaluated in light of the potential limitations of the present study. In discussing these limitations, we also suggest useful paths for future investigations. First, participants for this study were recruited from one university from each country and only two countries were represented, which may lead to concerns about the generalizability of these findings. Past research has suggested that equivalence is not a property unique to a measure. Instead, it is influenced by a host of factors, such as the characteristics of the sample and the context wherein the scale is administered (Robert et al., 2006). Therefore, it is recommended that additional research, which may include more countries and regions and multiple samples from each of them, be conducted to replicate these findings.

Second, the use of student participants may raise concerns about the external validity of our findings. Several factors should render this concern less tenable. First, there is some evidence suggesting that the results from studies using undergraduate samples can be generalized to field settings (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). Second, most of the participants had prior job application experiences and would be looking for a full-time job in the near future. Thus, these characteristics render them quite similar to actual job applicants. Third, in a post-experiment informal survey, most of the participants expressed strong motivations to do well in the faking condition. Nevertheless, future research should replicate our findings with a group of respondents who are actually applying for a job.

Conclusion

The present study established measurement equivalence of the BIDR across two cultural values: HI and HC. However, we found less support for the equivalence of the BIDR across two countries: the United States and Singapore, and across two motivational conditions: standard and faking. These results suggest that researchers should be particularly cautious when using the BIDR in cross-national contexts and when comparing the BIDR scores across groups varying in their motivations to engage in social desirability. These results also underscore the need to make measurement equivalence a routine procedure to provide the basis for any meaningful statistical analyses.

References

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81*, 261-272.

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107*, 238-246.

Bernardi, R. A. (2006). Association between Hofstede's cultural constructs and social desirability response bias. *Journal of Business Ethics, 65*, 43-53.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14,* 317-335.

Byrne, B. M. (2004). Testing for Multigroup Invariance Using *AMOS* Graphics: A Road Less Traveled. *Structural Equation Modeling, 11,* 272-300.

Byrne, B. M. & Watkins, D. (2003). The issue of measurement invariance revisit. *Journal of Cross-Cultural Psychology*, 34, 155-175.

Carmines, E. G., & McIver, J. P. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. W. Bohrnstedt & E. F. Borgatta (Eds.), *Social measurement: current issues* (65-115). Newbury Park, CA: Sage.

Chan, D. (2000). Detection of differential item functioning on the Kirton Adaptation-Innovation Inventory using multiple-group mean and covariance structure analyses. *Multivariate Behavioral Research, 35,* 169-199.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1-27.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31, 187-212.

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

Edwards, A. L. (1957). *The social desriabliyt variable in personality assessment and research.* New York: Dryden.

Ellingson, J. E., Sackett, O, R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84,* 155-166.

Ellingson, J. E., Smith, D. B., & Sackett, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology, 86,* 122-133.

Fox, S., & Schwartz, D. (2002). Social desirability and controllability in computerized and paper-and-pencil personality questionnaires. *Computer in Human Behavior, 18,* 389-410.

Geiger, M. A., & O'Connell, B. T. (2000). An examination of using surrogate measures to assess social desirability response bias. *Research on Accounting Ethics, 6,* 107-127.

Ghorpade, J., Hattrup, K., & Lackritz, J. R. (1999). The use of personality measures in cross-cultural research: A test of three personality scales across two countries. *Journal of Applied Psychology, 84*, 670-679.

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizational across nations.* Thousand Oaks, CA: Sage.

Holden, R. R. (1997). *Studying faking on personality tests: Statistical and research design issues*. Paper presented at the annual convention of the Canadian Psychological

Association, Toronto, Ontario, Canada.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

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

Lalwani, A. K., Shavitt, S., & Johnson, T. (2006). What is the relation between cultural orientation and socially desirable responding? *Journal of Personality and Social Psychology, 90,* 165-178.

Lam, S. S. K., Schaubroeck, J., & Aryee, S. (2002). Relationship between organizational justice and employee work outcomes: A cross-national study. *Journal of Organizational Behavior, 23,* 1-18.

Lay, C., Fairlie, P., Jackson, S., Ricci, T., Eisenberg, J., Sato, T. et al. (1998). Domain-specific allocentrism-idiocentrism: A measure of family connectedness. *Journal of Cross-Cultural Psychology, 29,* 434-460.

Li, A., & Bagger, J. (2007). The Balanced Inventory of Desirable Responding (BIDR): A Reliability Generalization Study. *Educational and Psychological Measurement*, *67*, 525-544.

Li, A., & Bagger, J. (2006). Using the BIDR to distinguish the effects of impression management and self-deception on the criterion validity of personality measures: A meta-analysis. *International Journal of Selection and Assessment, 14,* 131-141.

Little, T. D. (1997). Mean and covariance structures (MACS) analyses of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research, 32,* 53-76.

Liu, C., Borg, I., & Spector, P. E. (2004). Measurement equivalence of the German ob Satisfaction Survey used in a multinational organization: Implications of Schwartz's culture model. *Journal of Applied Psychology, 89,* 1070-1082.

Markus, H. R. & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224-253.

Mueller-Hanson, R., Heggestad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88,* 348-355.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychology attitudes* (pp. 17-59). New York: Academic Press.

Robert, C., Lee, W. C., & Chan, K-Y. (2006). An empirical analysis of measurement equivalence with the INDCOL measure of individualism and collectivism: Implications for valid cross-cultural inference. *Personnel Psychology, 59,* 65-99.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644.

Ryan, A. M., Chan, D., Ployhart, R. E., & Slade, L. A. (1999). Employee attitude surveys in a multinational organization: Considering language and culture in assessing measurement equivalence. *Personnel Psychology, 52*, 37-58.

Singelis, T. M., Triandis, H. C., Bhawuk, D., & Gelfand, M. J. (1995). Horizontal and vertical dimensions of individualism and collectivism: A theoretical and measurement refinement. *Cross-Cultural Research: The Journal of Comparative Social Science, 29*, 240-275.

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

Stober, J. (2001). The social desirability scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*, 222-232.

Triandis, H. C., Carnevale, P., Gelfand, M. J., Robert, C., Wasti, S. A., Probst, T. et al., (2001). Culture and deception in business negotiation: A multilevel analysis. *International Journal of Cross-Cultural Management, 1,* 73-90.

Triandis, H. C., & Gelfand, M. J. (1998). Converging measurement of horizontal and vertical individualism and collectivism. *Journal of Personality and Social Psychology, 74*, 118-128.

Triandis, H. C., & Suh, E. M. (2001). Cultural influences on personality. *Annual Review of Psychology, 53*, 133-160.

Tsui, A. S., Nifadkar, S. S., & Ou, A. Y. (2006). Cross-national, cross-cultural organizational behavior research: Advances, gaps, and recommendations. *Journal of Management, 33*, 426-478.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3,* 4-69.

Table 1: Demographic Information for Both Samples

	United States	Singapore
Sample Size	166	158
Mean Age	21	21
Percentage of Women	52.3%	54%
Percentage of Ethnic Majority	80%	83%

Table 2: Means, Standard Deviations, Intercorrelations, and Reliability Coefficients of Study Variables (American Sample)

	Mean	SD	SDE-S	IM-S	SDE-F	IM-F	HI	HC
SDE-S	4.24	.56	.63					
IM-S	3.61	.83	.50**	.82				
SDE-F	5.44	.59	.31**	.19*	.74			
IM-F	5.58	.89	.16*	.31**	.70**	.89		
HI	5.42	.58	.26**	.08	.21**	.19*	.54	
HC	5.34	.65	.02	.17*	-.01	-.01	.11	.67

Note. N = 166. SDE-S and IM-S are responses in the standard response condition; SDE-F and IM-F are responses in the faking response condition. Reliabilities are on the diagonal.

* $p < .05$

** $p < .01$

Table 3: Means, Standard Deviations, Intercorrelations, and Reliability Coefficients of Study Variables (Singaporean Sample)

	Mean	SD	SDE-S	IM-S	SDE-F	IM-F	HI	HC
SDE-S	3.95	.63	.74					
IM-S	3.46	.73	.29**	.79				
SDE-F	5.08	.73	.29**	.10	.82			
IM-F	4.73	1.02	-.03	.26**	.65**	.89		
HI	5.38	.59	.25**	.09	.09	-.09	.53	
HC	5.32	.68	-.03	.13	-.06	-.07	.27**	.72

Note. N = 158. SDE-S and IM-S are responses in the standard response condition; SDE-F and IM-F are responses in the faking response condition. Reliabilities are on the diagonal.

* $p < .05$

** $p < .01$

Table 4: Model Fit Indexes for All Models (BIDR: Two-Factor Model), Cross-Nations Analysis

	χ^2	*df*	χ^2/df	CFI	NFI	RMSEA	SRMR
Standard							
Model 1: American only	1243.25	739	1.68	.963	.914	.064	.083
Model 2: Singaporean only	1219.94	739	1.65	.963	.911	.064	.089
Model 3: baseline	2463.19	1478	1.67	.963	.912	.045	.083
Model 4: slope	2544.86	1516	1.68	.961	.910	.046	.091
Model 5: slope & intercept	2823.94	1554	1.82	.952	.900	.050	.092
Model 6: slope, intercept, & error variance	2919.34	1592	1.83	.950	.896	.051	.092
Faking							
Model 1: American only	1081.16	739	1.46	.983	.947	.053	.067
Model 2: Singaporean only	1157.14	739	1.57	.974	.932	.060	.074
Model 3: baseline	2238.32	1478	1.51	.979	.941	.040	.067
Model 4: slope	2299.15	1516	1.52	.978	.939	.040	.075
Model 5: slope & intercept	2597.36	1554	1.67	.971	.931	.046	.081
Model 6: slope, intercept, & error variance	2827.25	1592	1.78	.966	.925	.049	.081

$N_{Singapore} = 158$; $N_{U.S.} = 166$

Table 5: Model Fit Indexes for Horizontal Individualism and Horizontal Collectivism

	χ^2	*df*	χ^2/df	CFI	NFI	RMSEA	SRMR
HI							
Model 1: American only	30.10	20	1.51	.998	.993	.055	.062
Model 2: Singaporean only	27.53	20	1.38	.998	.993	.049	.063
Model 3: baseline	57.64	40	1.44	.998	.993	.037	.062
Model 4: slope	67.93	47	1.45	.997	.992	.037	.069
Model 5: slope & intercept	104.06	54	1.93	.994	.987	.054	.069
Model 6: slope, intercept, & error variance	118.04	61	1.94	.993	.986	.054	.073
HC							
Model 1: American only	31.95	20	1.60	.997	.993	.060	.054
Model 2: Singaporean only	23.52	20	1.18	.999	.994	.034	.048
Model 3: baseline	55.49	40	1.39	.998	.993	.035	.054
Model 4: slope	63.14	47	1.34	.998	.993	.033	.065
Model 5: slope & intercept	105.39	54	1.95	.994	.988	.054	.066
Model 6: slope, intercept, & error variance	117.42	61	1.93	.993	.986	.054	.065

$N_{Singapore} = 158$; $N_{U.S.} = 166$

Table 6: Model Fit Indexes for All Models (BIDR: Two-Factor Model), Cross-Cultures (Horizontal Individualism) Analysis

	χ^2	*df*	χ^2/df	CFI	NFI	RMSEA	SRMR
Standard							
Model 1: High HI only	1220.03	739	1.65	.963	.911	.064	.090
Model 2: Low HI only	1212.38	739	1.64	.965	.915	.063	.085
Model 3: baseline	2432.40	1478	1.65	.964	.913	.045	.090
Model 4: slope	2482.68	1516	1.64	.963	.911	.045	.094
Model 5: slope & intercept	2614.86	1554	1.68	.960	.907	.046	.095
Model 6: slope, intercept, error variance	2696.88	1592	1.69	.958	.904	.046	.094
Faking							
Model 1: High HI only	1085.10	739	1.47	.980	.941	.054	.070
Model 2: Low HI only	1228.09	739	1.66	.973	.935	.064	.071
Model 3: baseline	2313.19	1478	1.57	.977	.938	.042	.070
Model 4: slope	2355.54	1516	1.55	.976	.937	.041	.076
Model 5: slope & intercept	2393.41	1554	1.54	.976	.936	.041	.076
Model 6: slope, intercept, error variance	2462.41	1592	1.55	.976	.934	.041	.078

$N_{\text{High HI}} = 162$; $N_{\text{Low HI}} = 162$

Table 7: Model Fit Indexes for All Models (BIDR: Two-Factor Model), Cross-Cultures (Horizontal Collectivism) Analysis

	χ^2	*df*	χ^2/df	CFI	NFI	RMSEA	SRMR
Standard							
Model 1: High HC only	1230.75	739	1.67	.962	.910	.064	.088
Model 2: Low HC only	1267.04	739	1.72	.961	.911	.067	.089
Model 3: baseline	2497.79	1478	1.69	.961	.911	.046	.088
Model 4: slope	2541.46	1516	1.68	.961	.909	.046	.092
Model 5: slope & intercept	2614.36	1554	1.68	.960	.907	.046	.092
Model 6: slope, intercept, error variance	2684.59	1592	1.69	.959	.904	.046	.092
Faking							
Model 1: High HC only	1163.79	739	1.58	.976	.936	.060	.072
Model 2: Low HC only	1185.71	739	1.60	.976	.938	.061	.069
Model 3: baseline	2349.50	1478	1.59	.976	.937	.043	.072
Model 4: slope	2416.04	1516	1.59	.975	.935	.043	.081
Model 5: slope & intercept	2458.67	1554	1.58	.975	.934	.043	.081
Model 6: slope, intercept, error variance	2553.43	1592	1.60	.973	.932	.043	.082

$N_{High\ HC} = 162$; $N_{Low\ HC} = 162$

Table 8: Model Fit Indexes for All Models (BIDR: Two-Factor Model), Cross-Conditions Analysis

	χ^2	*df*	χ^2/df	CFI	NFI	RMSEA	SRMR
United States							
Model 1: Standard	1243.25	739	1.68	.963	.914	.064	.083
Model 2: Faking	1081.16	739	1.46	.983	.947	.053	.067
Model 3: baseline	2324.40	1478	1.57	.975	.933	.042	.083
Model 4: slope	2418.42	1516	1.60	.973	.931	.042	.093
Model 5: slope & intercept	2865.35	1554	1.84	.961	.918	.051	.101
Model 6: slope, intercept, error variance	3555.36	1592	2.23	.941	.898	.061	.104
Singapore							
Model 1: Standard	1219.94	739	1.65	.963	.911	.064	.089
Model 2: Faking	1157.14	739	1.57	.974	.932	.060	.074
Model 3: baseline	2377.08	1478	1.61	.969	.923	.044	.089
Model 4: slope	2518.00	1516	1.66	.966	.918	.046	.100
Model 5: slope & intercept	2800.71	1554	1.80	.957	.909	.051	.103
Model 6: slope, intercept, error variance	2987.52	1592	1.88	.952	.903	.053	.104

$N_{Singapore}$ = 158; $N_{U.S.}$ = 166