PsyCh Journal

INSTITUTE OF PSYCHOLOGY
Chinese Academy of Sciences

PsyCh Journal **5** (2016): 92–100
DOI: 10.1002/pchj.124

Measuring social desirability across language and sex: A comparison of Marlowe–Crowne Social Desirability Scale factor structures in English and Mandarin Chinese in Malaysia

A. Solomon Kurz,[1] Christopher F. Drescher,[2†] Eu Gene Chin[3], and Laura R. Johnson[1]

[1]Department of Psychology, University of Mississippi University, Mississippi, USA, [2]Department of Psychiatry and Health Behavior, Georgia Regents University, Augusta, Georgia, USA, [3]Department of Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, Mississippi, USA

Abstract: Malaysia is a Southeast Asian country in which multiple languages are prominently spoken, including English and Mandarin Chinese. As psychological science continues to develop within Malaysia, there is a need for psychometrically sound instruments that measure psychological phenomena in multiple languages. For example, assessment tools for measuring social desirability could be a useful addition in psychological assessments and research studies in a Malaysian context. This study examined the psychometric performance of the English and Mandarin Chinese versions of the Marlowe–Crowne Social Desirability Scale when used in Malaysia. Two hundred and eighty-three students (64% female; 83% Chinese, 9% Indian) from two college campuses completed the Marlowe–Crowne Social Desirability Scale in their language of choice (i.e., English or Mandarin Chinese). Proposed factor structures were compared with confirmatory factor analysis, and multiple indicators–multiple causes models were used to examine measurement invariance across language and sex. Factor analyses supported a two-factor structure (i.e., Attribution and Denial) for the measure. Invariance tests revealed the scale was invariant by sex, indicating that social desirability can be interpreted similarly across sex. The scale was partially invariant by language version, with some non-invariance observed within the Denial factor. Non-invariance may be related to differences in the English and Mandarin Chinese languages, as well as cultural differences. Directions for further research include examining the measurement of social desirability in other contexts where both English and Mandarin Chinese are spoken (i.e., China) and further examining the causes of non-invariance on specific items.

Keywords: factor analysis, language differences, Malaysia, sex differences, social desirability

Correspondence: A. Solomon Kurz, Department of Psychology, University of Mississippi, 205 Peabody, University, Mississippi 38677, USA. Email: ajkurz@go.olemiss.edu

Received 2 August 2015. Accepted 14 January 2016.

†Dr. Drescher and Dr. Chin began working on this project as graduate students in the Department of Psychology at the University of Mississippi.

Malaysia is a developing country in Southeast Asia where both English and Mandarin Chinese are prominently spoken among ethnic Chinese, who compose nearly one-quarter of the overall population (Central Intelligence Agency, 2013). Recently, psychological researchers have become interested in how differences in language versions of self-report questionnaires may affect responding among Malaysians (Chin et al., 2015), while others have noted the effects of sex on measurement (e.g., Burns, Walsh, Gomez, & Hafetz, 2006). The current study seeks to examine measurement invariance of a common measure of social desirability across language and sex in Malaysian college students.

Sources of bias

The behavioral sciences have historically focused on participants from Western, educated, industrialized, rich, and democratic countries (Henrich, Heine, & Norenzayan, 2010). Consequently, most studies in the existing literature focus on English-speaking samples. A recent analysis of publications from the 73 countries that comprise the International Union of Psychological Sciences for the years 2006–2010 found that 63% of psychological studies originate from English-speaking countries (O'Gorman, Shum, Halford, & Ogilvie, 2012). In contrast to their dominance in the psychological literature, English-speaking countries account for only 6% of the world's population (Henrich et al., 2010). One potential barrier to increasing research in non-English-speaking countries is the lack of empirically validated measures translated into other languages.

Extreme care must be taken when extending measurement methods developed in Western countries to other parts of the world (Fabri, 2008). When translating measures, investigators should address various types of equivalence between different language versions (Brislin, 1970; Fabri, 2008). One important type of equivalence is measurement equivalence: the "extent to which the psychometric properties of different language versions of the same instrument are similar" (Herdman, Fox-Rushby, & Badia, 1998, p. 330). This form of equivalence can be difficult to obtain when translating measures from English to Mandarin Chinese where differences in grammar and syntax and in concepts and word usage complicate translations (Yu, Lee, & Woo, 2004). In addition to examining measurement equivalence across languages, researchers should investigate potential differences in measurement properties across sex, given that sex differences present across numerous psychological phenomena and interact dynamically with culture (Guimond et al., 2007).

Social desirability

Social desirability is conceptualized as both response style on self-report measures and a facet of personality characterized by "the need of subjects to respond in culturally sanctioned ways" (Crowne & Marlowe, 1960, p. 354). From a research perspective, the measurement of social desirability may be useful to adjust for participants engaging in impression management and for detecting bias in many other domains, such as physical health complaints (Gravdal & Sandal, 2006). A widely used measure for assessing social desirability is the Marlowe–Crowne Social Desirability Scale (MCSDS; Crowne & Marlowe, 1960).

The conceptualization and measurement of social desirability is likely affected by cultural variables, and some studies suggest that social desirability may be higher among Asian persons (Middleton & Jones, 2000). However, conclusions from previous studies are limited by the language of self-report questionnaires. For example, though Middleton and Jones found that Asian college students in the US scored higher on the MCSDS than US college students, all questionnaires were completed in English, likely the second or third language for the Asian students.

As the original English scale was rather lengthy (33 dichotomous items), Ballard (1992) developed a 13-item English short-form composite. Although other short-forms have been developed, some researchers continue to recommend Ballard's composite due to its structural validity (Loo & Loewen, 2004). Another study demonstrated support for the structural validity of a Mandarin Chinese version of a 14-item short-form of the MCSDS in college students (Tao, Guoying, & Brody, 2009). This literature would benefit if Tao and colleagues' Chinese version of the MCSDS was replicated and extended in new cultural contexts, along with comparisons with English-language versions of the MCSDS.

Hypotheses

For the present study, we proposed three key hypotheses. First, considering previous research (Tao et al., 2009), we expected a two-factor model to demonstrate acceptable fit and to be superior to a one-factor model. Second, we expected differences between language versions (Herdman et al., 1998) resulting in only partial measurement invariance between them. However, we had no a priori expectations about which parameters might perform non-invariantly. Third, given previous research (e.g., Bobbio & Manganelli, 2011), we expected full measurement invariance by sex.

Method

Participants

Participants were a diverse sample of college students recruited from INTI International College Subang and Methodist College Kuala Lumpur. Both colleges are private educational institutions located in Klang Valley, a region widely considered an educational hub on the western side of the Malaysian Peninsula. INTI International College

Subang is located in Subang Jaya, a suburban city, and Methodist College Kuala Lumpur is located in Kuala Lumpur, the capital of Malaysia.

Measure: MCSDS, short-form

We measured social desirability using Tao et al.'s (2009) 14-item short-form of the MCSDS. Tao and colleagues reported that they translated one of the short-forms presented by Ballard (1992) into Chinese. In addition to reviewing the factor structures of short-forms presented by previous authors, Ballard presented four alternative short-forms. Comparison of the items included in Ballard's short-forms with those included by Tao and colleagues revealed that Tao and colleagues' version contained all of the 13 items in Ballard's "Composite" (i.e., MCSDS Items 6, 10, 12–16, 19, 21, 26, 28, 30, and 33) with the addition of MCSDS Item 11, "I like to gossip at times." Thus, we compared both models in the present study.

Procedures

The current study is an analysis of MCSDS data from a larger cross-sectional study (see Chin et al., 2015) approved by the Institutional Review Board at the authors' home university, the University of Mississippi, as well as the administrative bodies at Methodist College Kuala Lumpur and INTI International College Subang. On both campuses, students were recruited with in-class announcements. After informed consent was obtained via procedures consistent with American Psychological Association (2010) ethical principles, participants completed the self-report measures in their language of choice, Chinese or English.

Statistical analyses

Data preparation

Mahalanobis D^2 values indicated no multivariate outliers. Missing values rates were 2.1% for the Chinese version and 1.3% for the English version. In order to maximize power for our modest sample sizes and accommodate the categorical nature of our variables, we handled missing values with multiple imputation, under the missing at random assumption, and via the sequential regression approach, a Bayesian procedure shown to perform particularly well for non-normal distributions (Van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006). Using M*plus* 7.11 (L. K. Muthén & Muthén, 2012), we generated 100 imputed datasets (see Enders, 2010) using an unrestricted model, including auxiliary variables from the full dataset.

Confirmatory factor analysis

To accommodate the dichotomous items of the MCSDS and to utilize fit indices, we performed confirmatory factor analyses (CFA) using the non-normality robust weighted least square estimator (i.e., WLSMV) in M*plus*. We set latent variable metrics by constraining factor variances to 1 in all models. In addition to the model χ^2, we assessed model fit with the root mean square error of approximation (RMSEA; acceptable < .08, good fit < .05), the comparative fit index (CFI; acceptable > .90, good fit > .95), and the Tucker–Lewis index (TLI; acceptable > .90, good fit > .95; Brown, 2015). With multiple imputation, model fit indices are pooled across the imputed datasets. As methodologists have yet to establish how to best pool the indices (Enders, 2010) or to best pool the likelihood ratio test for WLSMV (Asparouhov & Muthén, 2010a), we followed Enders' (2010) ad hoc procedure of presenting the mean and distribution (i.e., 95% confidence intervals [CI]) of the pooled fit indices. Also, it is not currently possible to use the M*plus* DIFFTEST option to perform likelihood ratio tests when analyzing multiply imputed datasets with the WLSMV estimator (L. K. Muthén, 2012). In the absence of likelihood ratio testing, we assessed model comparisons by noting the pattern of numerical changes in the χ^2 and other fit indices, across all models.

Multiple indicators–multiple causes modeling

After estimating the simple CFA measurement models, we fit multiple indicators–multiple causes (MIMIC; a.k.a., CFA with covariates; B. O. Muthén, 1989) models to examine measurement invariance for language and sex. We elected this method over the more frequently used multigroup CFA (MGCFA) method because of our modest sample size ($N = 283$). That is, with the MGCFA method, the input matrix is divided by groups, factor models are estimated separately for each group, and parameter classes are sequentially constrained to equality across groups (see Brown, 2015). For the present dataset, that would entail splitting the input matrix into $n = 155$ and $n = 128$ for MGCFA by language and $n = 181$ and $n = 102$ for MGCFA by sex. With MIMIC models, however, invariance across groups is tested with a single input matrix, granting invariance tests the power and the parameter stability associated with the full sample size. With MIMIC models for categorical data, measurement invariance may be examined in terms of: (a) differing factor means; and (b) differing item thresholds (also known as differential item functioning; e.g., Song, Cai, Brown, & Grimm, 2011). Procedurally, we first examined invariance of latent variable means by regressing the factors on covariates of interest. Our second step involved examining direct effects of the items on the covariates, with the effects of the covariates on the factors held constant. Significant direct effects on factors and/or item intercepts indicated invariance.

Results

Demographics

Two hundred and eighty-six participants consented to the study. We omitted data from three participants—all of whom chose the Chinese battery—who discontinued participation after partially completing the demographic questions. Of those who completed the MCSDS in Chinese (n = 128; Mdn_{age} = 18.0, SD = 1.0) 64% identified as female, and in terms of ethnicity, 95% identified as Chinese and 2% as Indian. Of those who completed the questionnaires in English (n = 155; Mdn_{age} = 18.0, SD = 0.9) 64% identified as female, while for ethnicity, 73% identified as Chinese, 14% as Indian, 3% as Malay, and 10% as Other.

Confirmatory factor analyses

In order to replicate the findings of Tao et al. (2009), we first compared a one- and two-factor structure for the full 14-item measure. See Table 1 for descriptive statistics and correlations for the MCSDS items, sex, and language. Similar to Tao and colleagues' results, the fit indices indicated that the two-factor solution (e.g., $\text{RMSEA}_{\text{pooled}}$ = .035) was preferable to the one-factor solution (e.g., $\text{RMSEA}_{\text{pooled}}$ = .046; see Table 2 for fit indices for all models). In order to compare these results with the original 13-item Ballard Composite (1992), we omitted MCSDS Item 11. Similar to the results of Tao and colleagues' 14-item model, the two-factor solution for Ballard's Composite (e.g., $\text{RMSEA}_{\text{pooled}}$ = .036) showed preferable fit[1] when compared to the one-factor solution (e.g., $\text{RMSEA}_{\text{pooled}}$ = .049). Our initial comparison of the fit indices of the two-factor versions for Ballard's Composite and Tao and colleagues' model led us to conclude that there was little statistical support for one over the other.

Close inspection of the item loadings for all four models revealed that MCSDS Item 19 ("I sometimes try to get even rather than to forgive and forget") loaded non-significantly in all models (see Table 3). After re-estimating all four models with the omission of that item, the two-factor solutions still showed better fit than the one-factor models for both Ballard's Composite and Tao and colleagues' model. All items for these models loaded significantly, with the exception of MCSDS Item 11 for Tao and colleagues'$_{-19}$ model, which contained zero within the bounds of its CIs, λ = 0.22, 95% CI [0.00, 0.45]. This was

Table 1
Category Proportions and Tetrachoric Correlations between the MCSDS, Language, and Sex

	Category	Tetrachoric correlations														
Variable	0/1	6	10	11	12	13	14	15	16	19	21	26	28	30	33	Lang
MCSDS item																
6	.74/.26															
10	.66/.34	.37														
11	.80/.21	.25	.21													
12	.50/.50	.03	.08	−.09												
13	.81/.19	−.18	−.20	−.16	−.11											
14	.64/.36	.20	.01	.22	.20	−.21										
15	.66/.34	.18	.06	.09	.44	−.28	.26									
16	.60/.40	−.21	.01	.00	−.21	.33	−.07	−.30								
19	.60/.40	.17	.17	.15	.05	−.01	−.15	.04	.00							
21	.68/.32	−.18	−.21	−.06	.03	.28	.00	−.24	.05	−.14						
26	.49/.51	−.13	−.03	−.17	.06	.45	.09	−.13	.14	−.27	.24					
28	.75/.25	.30	.15	−.02	.32	−.16	.18	.27	−.09	−.04	−.19	−.20				
30	.47/.53	.19	.21	.10	.28	−.20	.10	.40	.01	.13	−.27	−.10	.37			
33	.52/.48	−.02	.03	−.10	−.11	.30	−.07	−.36	.24	−.03	.22	.25	−.04	−.01		
Language	.45/.55	−.23	−.10	.26	−.33	.16	−.18	−.24	.06	.15	.18	−.06	−.43	−.38	.09	
Sex	.68/.32	−.01	−.05	.08	−.05	−.06	.13	−.24	.08	.09	.02	−.02	−.16	−.06	.08	.02

Note. Category = the proportions of respondents endorsing the first (i.e., 0) or second (i.e., 1) response option for the dichotomous variables (MCSDS coded 0 = no, 1 = yes; Language coded 0 = Mandarin Chinese, 1 = English; Sex coded 0 = female, 1 = male); MCSDS = Marlow–Crowne Social Desirability Scale. As missing data were handled using multiple imputation, the proportions are average results over the multiply imputed datasets.

Table 2
Fit Statistics for the MCSDS CFA and MIMIC Models

Model	χ_{pooled}^2 *M*	χ_{pooled}^2 95% CI	*df*	$RMSEA_{pooled}$ *M*	$RMSEA_{pooled}$ 95% CI	CFI_{pooled} *M*	CFI_{pooled} 95% CI	TLI_{pooled} *M*	TLI_{pooled} 95% CI	λ^{ns}	Invariance for Factor means	Invariance for Item thresholds
CFA models												
Tao and colleagues'												
1-factor	123.250	[123.071, 123.429]	77	.046	[.046, .046]	.795	[.794, .796]	.757	[.756, .758]	19	NA	NA
2-factor	102.911	[102.710, 103.112]	76	.035	[.035, .035]	.880	[.879, .881]	.857	[.856, .858]	19	NA	NA
2-factor$_{-19}$	83.499	[83.375, 83.623]	64	.033	[.033, .033]	.911	[.910, .912]	.892	[.892, .892]	11	NA	NA
Ballard's composite												
1-factor	109.202	[109.037, 109.367]	65	.049	[.049, .049]	.801	[.800, .802]	.761	[.761, .761]	19	NA	NA
2-factor	87.621	[87.433, 87.809]	64	.036	[.036, .036]	.893	[.892, .984]	.870	[.869, .871]	19	NA	NA
2-factor$_{-19}$	68.878	[68.765, 68.991]	53	.033	[.033, .033]	.927	[.926, .928]	.909	[.909, .909]		NA	NA
MIMIC models: Factors and items regressed on Language												
Tao and colleagues'												
2-factor	105.334	[105.165, 105.503]	85	.029	[.029, .029]	.917	[.916, .918]	.897	[.897, .897]		Denial	11 15[a] 19
2-factor	109.084	[108.916, 109.252]	86	.031	[.031, .031]	.905	[.904, .906]	.885	[.885, .885]		Denial	11 19
2-factor$_{-19}$	87.311	[87.198, 87.424]	73	.026	[.026, .026]	.939	[.938, .940]	.924	[.924, .924]		Denial	11 15
Ballard's Composite												
2-factor	9.773	[9.608, 9.938]	73	.029	[.029, .029]	.923	[.922, .924]	.923	[.923, .923]		Denial	15 19
2-factor$_{-19}$	72.467	[72.362, 72.572]	62	.024	[.024, .024]	.953	[.953, .953]	.941	[.941, .941]		Denial	15
MIMIC models: Factors regressed on Sex												
Tao and colleagues'												
2-factor	113.849	[113.689, 114.009]	88	.032	[.032, .032]	.885	[.884, .886]	.863	[.862, .864]	19		NA
2-factor$_{-19}$	92.862	[92.632, 93.092]	75	.029	[.029, .029]	.919	[.918, .920]	.901	[.900, .902]	11		NA
Ballard's Composite												
2-factor	98.043	[97.772, 98.314]	75	.033	[.033, .033]	.896	[.895, .897]	.874	[.873, .875]	19		NA
2-factor$_{-19}$	77.604	[77.356, 77.852]	63	.029	[.029, .029]	.932	[.931, .933]	.916	[.915, .917]			NA

Note. λ^{ns} = MCSDS number for items with non-significant factor loadings; CFA = confirmatory factor analyses; CFI = comparative fit index; CI = confidence interval; MCSDS = Marlow–Crowne Social Desirability Scale; MIMIC = multiple indicators–multiple causes; NA = not applicable; RMSEA = root mean square error of approximation; TLI = Tucker–Lewis index.

[a] The 95% CIs for the direct effect of Language on this item just contained zero within their bounds [.00, .45] in this model.

Table 3
Factor Loadings and Factor Correlations Based on the Unconditional Confirmatory Factor Analyses for the Four Two-Factor Models

Item #		Tao and colleagues		Tao and colleagues$_{-19}$		Ballard's Composite		Ballard's Composite$_{-19}$	
Tao	MC	Est.	95% CIs	Est.	95% CIs	Est.	95% CIs	Est.	95% CIs
				Attachment factor loadings (λ)					
5.	13.	**.74**	[.52, .97]	**.76**	[.52, .99]	**.74**	[.51, .97]	**.75**	[.52, .99]
8.	16.	**.40**	[.21, .60]	**.41**	[.21, .61]	**.41**	[.21, .61]	**.42**	[.22, .61]
1.	21.	**.46**	[.26, .66]	**.45**	[.25, .65]	**.46**	[.26, .66]	**.45**	[.25, .65]
11.	26.	**.48**	[.30, .66]	**.47**	[.29, .65]	**.48**	[.30, .66]	**.46**	[.28, .65]
14.	33.	**.46**	[.28, .63]	**.46**	[.28, .64]	**.46**	[.28, .64]	**.46**	[.28, .64]
				Denial factor loadings (λ)					
1.	6.	**.45**	[.26, .63]	**.43**	[.25, .62]	**.42**	[.24, .61]	**.41**	[.22, .60]
2.	10.	**.32**	[.14, .50]	**.31**	[.12, .49]	**.30**	[.12, .48]	**.29**	[.10, .47]
3.	11.	**.24**	[.01, .46]	.22	[.00, .45]				
4.	12.	**.44**	[.26, .61]	**.45**	[.27, .62]	**.46**	[.29, .64]	**.47**	[.29, .65]
6.	14.	**.30**	[.12, .48]	**.32**	[.14, .50]	**.29**	[.11, .47]	**.31**	[.13, .49]
7.	15.	**.70**	[.54, .86]	**.72**	[.56, .88]	**.72**	[.56, .88]	**.73**	[.57, .89]
9.	19.	.16	[−.03, .35]			.15	[−.04, .34]		
12.	28.	**.52**	[.34, .70]	**.53**	[.35, .71]	**.53**	[.35, .71]	**.54**	[.36, .72]
13.	30.	**.55**	[.38, .72]	**.54**	[.37, .71]	**.55**	[.38, .72]	**.54**	[.37, .71]
				Factor correlation (ψ)					
		−.56	[−.75, −.36]	**−.54**	[−.74, −.35]	**−.54**	[−.74, −.35]	**−.53**	[−.73, −.34]

Note. CI = confidence interval; Tao = numbering used by Tao and colleagues; MC = numbering used by the original 33-item Marlowe–Crowne Social Desirability Scale; Est. = point estimate. Factor loadings and factor correlations with 95% CIs not containing zero within their bounds are in boldface.

noteworthy because this was the item that Tao and colleagues added to Ballard's Composite.

MIMIC models

Invariance by language

To test for measurement invariance based on questionnaire language, we first regressed the factors, Attribution and Denial, on the binary covariate "Language" (coded 0 = Chinese, 1 = English) for the four two-factor models presented in the CFA section. The four models retained adequate fit when regressed on Language. Next, we examined the direct effects of Language on the items, while controlling for the effects of Language on the factors. Results revealed a fairly consistent pattern of threshold non-invariance across three items: MCSDS Items 11, 15, and 19. The regression coefficients were positive in all cases, indicating that participants who completed the English version of the MCSDS had higher thresholds (i.e., higher probabilities of endorsing those items as "yes," with their overall latent scores held constant) than participants who completed the Chinese version. The one exception was for Tao and colleagues' model, for which the direct effect of language on MCSDS Item 15 was not significant, $b = 0.428$, 95% CI [−0.025, 0.881]. For the other models, including Item 15, the 95% CIs did not contain zero within their bounds. For all models, Language negatively and significantly predicted the Denial factor, but not the Attribution factor. Using Tao and colleagues' model as an example, this indicates that participants who completed the English version of the MCSDS exhibited lower levels of the Denial factor than their peers who completed the questionnaire in Chinese, $b = -1.21$, 95% CI [−1.69, −0.73]. As the factors were in a standardized metric, this meant participants who completed the English version scored more than a full standard deviation lower in Denial, on average, than those who completed the Chinese version—a rather large effect. However, results suggested there was no statistical difference between the two groups for the Attribution factor.

Invariance by sex

Next, to test if participants exhibited measurement invariance on the MCSDS by sex, we regressed the two-factor measurement models on the binary covariate "Male" (coded 0 = female, 1 = male). As with Language, we did this for all four of the two-factor models presented in the CFA section. The four models generally retained adequate fit when the factors were regressed on Male (see Table 2). Results revealed no significant direct effects for Male on any of the items or on either of the factors for any of the models. Thus, the MCSDS models herein were invariant by Male in terms of factor means and item thresholds.

Discussion

Although the MCSDS has the potential to assess social desirability (e.g., easy to use, brief, freely available), its psychometrics have not been previously evaluated with Malaysian participants. In this study, we examined the factor structure of two short-forms of the MCSDS across two language versions (Chinese and English) in a sample of Malaysian college students. As hypothesized, the two-factor structure (i.e., Attribution and Denial) was superior to a one-factor factor structure.

Also congruent with predictions, we found that three items had differing sensitivity levels across languages. Specifically, participants who completed the Chinese version had a higher probability for endorsing "no" to Items 11 ("I like to gossip at times"), 15 ("There have been occasions when I took advantage of someone"), and 19 ("I sometimes try to get even rather than to forgive and forget") compared to participants who completed the English version, given the same level of the Denial factor (i.e., the negatively worded factor). Additionally, comparison of factor means suggested that participants who completed the MCSDS in Chinese scored higher levels of the Denial factor than those who completed the measure in English. The strong norms of moderation (Cheung & Song, 1989) and stigma associated with distress (Chan & Parker, 2004) within Chinese culture may partially explain the higher levels of Denial for those who completed the MCSDS in Chinese. Completing the measure in Chinese may cue these norms more strongly than completing the measure in English.

Finally, as hypothesized, results indicated full measurement invariance for item thresholds and factor means across sex, suggesting that results can be interpreted similarly across males and females. Researchers and clinicians would not have to worry, for instance, that sex roles (Bobbio & Manganelli, 2011) would have any impact on MCSDS scores, enabling a fair comparison of male and female MCSDS scores.

In addition to the general findings concerning our hypotheses, two issues arose that merit further discussion. First, MCSDS Item 19 ("I sometimes try to get even rather than to forgive and forget") loaded non-significantly on all four-factor models examined in this study. This item has not been reported as troublesome in previous factor analyses of MCSDS short-forms in English or Chinese, suggesting that there may be something specific to the Malaysian participants of this study that affected its performance. Part of Item 19, "get even," may be considered an idiom (Bhui, Mohamud, Warfa, Craig, & Stansfeld, 2003) that is meaningless to Malaysian college students, even for those taking the measure in English. For instance, instead of precisely denoting the act of getting revenge, this term may denote alternative meanings, such as accepting the situation or smoothing out the situation. A systematic qualitative inquiry into this issue may better delineate the meaning of this term for Malaysians.

A second important finding is a trend towards the superiority of Ballard's (1992) 13-item composite over Tao et al.'s (2009) 14-item short-form of the MCSDS. Specifically, MCSDS Item 11 ("I like to gossip at times") performed poorly across a number of analyses. As Tao and colleagues presented no clear rationale in their paper for adding MCSDS Item 11 to Ballard's Composite, we recommend excluding the item in future studies.

Strengths and limitations

A strength for the present study is that it is the first to assess the factor structure of the MCSDS in the Malaysian context and directly compare the factor structures of Chinese and English versions. Understanding how the measure performs across language versions is essential given the variety of languages spoken across Malaysia. Moreover, psychometrically sound measures of social desirability would be useful for psychological assessments in Malaysia because mental health is still a highly stigmatized topic in Malaysia (e.g., Minas, Zamzam, Midin, & Cohen, 2011). Yet, because of our modest sample size and corresponding use of MIMIC models in place of MGCFA, we were not able to assess for equality across model variances and covariances. The ethnic composition across the two language versions was imbalanced; while 95% of participants taking the Chinese version of the measure identified ethnically as Chinese, only 73% of participants taking the English version identified as such. Therefore, differences across language versions may be, in part, attributable to cultural differences between the two samples. We did not assess for bilingualism, leaving its effects on response styles across languages uncertain (Gibbons, Zellner, & Rudek, 1999). Finally, this study is limited by the lack of follow-up qualitative inquiries of the problematic MCSDS items.

Future directions

The MCSDS literature would benefit from further comparisons between English and Chinese versions in other contexts where both languages are spoken (e.g., China, US).

The finding that MCSDS Item 19 did not significantly load on any factor needs replication before researchers exclude it from future studies. Additionally, future studies might address the potential effects of language on MCSDS Items 15 and 19. For example, researchers could examine language effects by conducting interviews with bilinguals to qualitatively examine the connotations and denotations elicited by different wordings for items. By working with larger samples, researchers might use the more rigorous MGCFA method to examine all possible sources of measurement invariance across groups. Lastly, researchers could also balance the ethnic composition across English and Chinese samples for a quantitative examination.

Conclusion

The current study gave further support for the superiority of the two-factor structure to the one-factor structure for the MCSDS in Malaysian college students across two language versions, Chinese and English. Our analyses indicated that the psychometric properties of the Attribution factor were the same for the Chinese and the English versions, allowing for direct comparisons between languages. In addition, the measurement was invariant by sex, making direct comparisons across sex feasible in Malaysian college students. However, participants who completed the measure in Chinese had a lower mean in the Denial factor and a lower probability to report "yes" to Items 11, 15, and 19, complicating direct comparisons between Malaysian individuals completing the MCSCS in Chinese and English. Future research may clarify these differences by employing qualitative methods (e.g., interviews, focus groups) to guide future item revisions and quantitative methods to further parse out the influence of bilingualism, method effect, and ethnic group.

Note

1. To supplement our analyses, we compared the one- and two-factor models using WLSMV estimation on the raw data, without imputation. In such a case, the WLSMV estimator is not technically equivalent to full information maximum likelihood estimation or multiple imputation, but does accommodate the missing values under the assumption of missing at random after controlling for covariates (Asparouhov & Muthén, 2010b). In order to make this assumption reasonable, we used sex, ethnicity, language version, and location as auxiliary variables. The overall model fit results for the one- and two-factor models using both Tao and colleagues' version and Ballard's Composite were comparable to those when using multiple imputation. The nested χ^2 difference tests revealed that the one-factor model fit was significantly worse than the two-factor model for both Tao and colleagues' version, $\chi^2(1) = 14.420$, $p = .0001$, and Ballard's Composite, $\chi^2(1) = 15.540$, $p = .0001$.

References

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Retrieved from http://www.apa.org/ethics/code/principles.pdf

Asparouhov, T., & Muthén, B. O. (2010a). *Chi-square statistics with multiple imputation* (Mplus Technical Appendix). Retrieved from http://www.statmodel.com/download/MI7.pdf

Asparouhov, T., & Muthén, B. O. (2010b). *Weighted least squares estimation with missing data* (Mplus Technical Appendix). Retrieved from http://www.statmodel.com/download/GstrucMissingRevision.pdf

Ballard, R. (1992). Short forms of the Marlowe-Crowne Social Desirability Scale. *Psychological Reports*, *71*(3, Pt 2), 1155–1160. doi:10.2466/pr0.1992.71.3f.1155

Bhui, K., Mohamud, S., Warfa, N., Craig, T. J., & Stansfeld, S. A. (2003). Cultural adaptation of mental health measures: Improving the quality of clinical practice and research. *British Journal of Psychiatry*, *183*, 184–186. doi:10.1192/bjp.183.3.184

Bobbio, A., & Manganelli, A. M. (2011). Measuring social desirability responding: A short version of Paulhus' BIDR 6. *TPM-Testing, Psychometrics, Methodology in Applied Psychology*, *18*, 117–135.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*, 185–216. doi:10.1177/135910457000100301

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.

Burns, G. L., Walsh, J. A., Gomez, R., & Hafetz, N. (2006). Measurement and structural invariance of parent ratings of ADHD and ODD symptoms across gender for American and Malaysian children. *Psychological Assessment*, *18*, 452–457. doi:10.1037/1040-3590.18.4.452

Central Intelligence Agency. (2013). *The world factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/

Chan, B., & Parker, G. (2004). Some recommendations to assess depression in Chinese people in Australasia. *Australian and New Zealand Journal of Psychiatry*, *38*, 141–147. doi:10.1080/j.1440-1614.2004.01321.x

Cheung, F. M., & Song, W. (1989). A review on the clinical applications of the Chinese MMPI. *Psychological Assessment*, *1*, 230–237. doi:10.1037/1040-3590.1.3.230

Chin, E. G., Drescher, C. F., Trent, L. R., Darden, M., Seak, W. C., & Johnson, L. R. (2015). Searching for a screener: Examination of the factor structure of the General Health Questionnaire in Malaysia. *International Perspectives in Psychology: Research, Practice, Consultation*, *4*, 111–127. doi:10.1037/ipp0000030

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354. doi:10.1037/h0047358

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Fabri, M. (2008). Cultural adaptation and translation of assessment instruments for diverse populations: The use of the Harvard Trauma Questionnaire in Rwanda. In L. A. Suzuki & J. G. Ponterotto (Eds.), *Handbook of multicultural assessment: Clinical, psychological and educational applications* (3rd ed., pp. 195–219). San Francisco, CA: John Wiley & Sons.

Gibbons, J. L., Zellner, J. A., & Rudek, D. J. (1999). Effects of language and meaningfulness on the use of extreme response style by Spanish-English bilinguals. *Cross-Cultural Research*, *33*, 369–381. doi:10.1177/106939719903300404

Gravdal, L., & Sandal, G. M. (2006). The two-factor model of social desirability: Relation to coping and defense, and implications for health. *Personality and Individual Differences*, *40*, 1051–1061. doi:10.1016/j.paid.2005.11.004

Guimond, S., Branscombe, N. R., Brunot, S., Buunk, A. P., Chatard, A., Désert, M., … Yzerbyt, V. (2007). Culture, gender, and the self: Variations and impact of social comparison processes. *Journal of Personality and Social Psychology*, *92*, 1118–1134. doi:10.1037/0022-3514.92.6.1118

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. doi:10.1017/S0140525X0999152X

Herdman, M., Fox-Rushby, J., & Badia, X. (1998). A model of equivalence in the cultural adaptation of HRQoL instruments: The universalist approach. *Quality of Life Research*, *7*, 323–335. doi:10.1023/A:1024985930536

Loo, R., & Loewen, P. (2004). Confirmatory factor analyses of scores from full and short versions of the Marlowe–Crowne Social Desirability Scale. *Journal of Applied Social Psychology*, *34*, 2343–2352. doi:10.1111/j.1559-1816.2004.tb01980.x

Middleton, K. L., & Jones, J. L. (2000). Socially desirable response sets: The impact of country culture. *Psychology and Marketing*, *17*, 149–163. doi:10.1002/(SICI)1520-6793(200002)17:2<149::AID-MAR6>3.0.CO;2-L

Minas, H., Zamzam, R., Midin, M., & Cohen, A. (2011). Attitudes of Malaysian general hospital staff towards patients with mental illness and diabetes. *BMC Public Health*, *11*, 317. doi:10.1186/1471-2458-11-317

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–585. Retrieved from https://www.statmodel.com/download/Article_0241.pdf

Muthén, L. K. (2012, June 6). *Mplus discussion > Confirmatory factor analysis > DIFFTEST*. Message posted to http://www.statmodel.com/discussion/messages/9/1865.html?1443657554

Muthén, L. K., & Muthén, B. O. (2012). Mplus user's guide (7th ed.). Los Angeles, CA: Muthén & Muthén.

O'Gorman, J., Shum, D. H. K., Halford, W. K., & Ogilvie, J. (2012). World trends in psychological research output and impact. *International Perspectives in Psychology: Research, Practice, Consultation*, *1*, 268–283. doi:10.1037/a0030520

Song, H., Cai, H., Brown, J. D., & Grimm, K. J. (2011). Differential item functioning of the Rosenberg Self-Esteem Scale in the US and China: Measurement bias matters. *Asian Journal of Social Psychology*, *14*, 176–188. doi:10.1111/j.1467-839X.2011.01347.x

Tao, P., Guoying, D., & Brody, S. (2009). Preliminary study of a Chinese language short form of the Marlowe–Crowne Social Desirability Scale. *Psychological Reports*, *105*, 1039–1046. doi:10.2466/PR0.105.F.1039-1046

Van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*, 1049–1064. doi:10.1080/10629360600810434

Yu, D. S. F., Lee, D. T. F., & Woo, J. (2004). Issues and challenges of instrument translation. *Western Journal of Nursing Research*, *26*, 307–320. doi:10.1177/0193945903260554