STANDARDIZATION TO ACCOUNT FOR CROSS-CULTURAL RESPONSE BIAS
A Classification of Score Adjustment Procedures and Review of Research in *JCCP*

RONALD FISCHER
Victoria University Wellington

The article reviews standardization methods commonly employed to adjust for response bias in cross-cultural research. First, different standardization procedures are reviewed and a classification scheme is provided. Standardization procedures are classified according to the statistical information used (means, standard deviation) and the source of this information (individual, group, or culture). Second, empirical research in *JCCP* between 1970 and 2002 is reviewed. Standardization has become more common in the 1990s, and there is a trend to rely more on standardized data. Most studies used standardization prior to analysis of variance and factor analytical techniques. However, an analysis of statistical properties of standardized measures indicates that results based on standardization are ambiguous. The use of statistical techniques and the interpretation of results based on standardized data are discussed.

Keywords: response bias; standardization; ipsative scores; cross-cultural research; factor analysis; score adjustment; cross-cultural differences

Cross-cultural researchers conducting studies with individuals from different cultural groups need to consider whether the scores obtained are comparable. Equivalence and bias are important issues that need to be addressed before meaningful cross-cultural comparisons can be made (Van de Vijver & Leung, 1997). Response bias is the systematic tendency to distort responses to rating scales so that observed scores are unrelated to the true score of the individual by either selecting extreme or modest answers (extreme or modesty response bias) or a shifting of responses to either end of the scale (acquiescence response bias) (Byrne & Campbell, 1999; Cheung & Rensvold, 2000). Hofstede (1980) was among the first to advocate the use of standardization as an adjustment of raw scores in cross-cultural research to correct for such response tendencies (p. 77ff.). The principal aim of standardization is a reduction or elimination of unwanted cross-cultural differences that are not due to variables of interest, but rather response sets and methodological artifacts (Van de Vijver & Leung, 1997). However, the term *standardization* has been used in a variety of contexts and different methods of standardization have emerged over the years. Therefore, the present article first provides a summary and classification of different standardization procedures. Second, it reviews research published in the *Journal of Cross-Cultural Psychology* (*JCCP*) between 1970 and 2002 to examine trends in the use of standardization procedures. Using the classifi-

AUTHOR'S NOTE: A previous version of this article was presented at the Regional Congress of the International Association for Cross-Cultural Psychology in Budapest, Hungary, July 2003. I would like to thank Dianne van Hemert, Peter B. Smith, David Matsumoto, Walter Lonner, one anonymous reviewer, Christopher Sibley, and members of the Cross-Cultural Research Group at Victoria University for their encouragement, helpful comments, and discussions. I would also like to thank Shaun Dugan and Charles Harb for many helpful and insightful discussions concerning response biases and standardizations over the years.

JOURNAL OF CROSS-CULTURAL PSYCHOLOGY, Vol. 35 No. 3, May 2004 263-282
DOI: 10.1177/0022022104264122

cation scheme, it is investigated which standardization procedures have been used, what the purpose of standardization was (e.g., analysis of mean differences, structural equivalence of measures), and the statistical techniques used with standardized scores. Third, the use of certain data analytical methods for standardized data is questioned by highlighting some undesirable properties of standardization. Based on these three issues, some recommendations for cross-cultural bias analysis and cross-cultural comparisons are made.

Before presenting the classification framework of standardization procedures, it should be pointed out that the purpose of this article is neither to discuss the nature of these response biases nor appropriate techniques to detect response patterns in cross-cultural data sets. Concerning the nature of response patterns, Hofstede (1980) and Van de Vijver and Leung (1997) argue that different response patterns are some form of method bias, and researchers need to detect and control for this bias or error variance in cross-cultural research. Assuming that different patterns are some form of bias, researchers might want to standardize their data to reduce this error variance. In contrast to this bias view, Smith (2004) and Van Hemert, Van de Vijver, Poortinga, and Georgas (2002) showed that these response patterns might not be bias per se but rather communication styles related to cultural characteristics. According to this view, standardization would remove variation that is substantial and related to culture. Therefore, standardization should not be used, and such differences in response styles should be studied in their own right as manifestations of culture. These two perspectives are not necessarily mutually exclusive. For example, cultural tendencies to respond in a more acquiescent style may reflect underlying cultural characteristics such as politeness or deference to high-status individuals (researchers or superiors administering questionnaires) or general social agreement tendencies. Nevertheless, these cultural tendencies are likely to change the responses of participants and make them incomparable across cultural groups, therefore resulting in a bias. Depending on the explicit and implicit focus of a research project, either focus might become more appropriate. This is clearly an area that needs to be investigated in more detail.

Concerning techniques detecting such patterns, Cheung and Rensvold (2000) and Little (1997, 2000) discussed strategies using structural equation modeling, whereas Van de Vijver and Leung (1997) discussed approaches using item response theory. Item response theory has been quite popular and various studies have employed this technique (for recent examples, see Butcher, Lim, & Nezami, 1998; Huang, Church, & Katigbak, 1997). These techniques can be used to detect whether different response strategies were used by participants in different samples. Researchers might then decide whether they want to standardize their data.

A CLASSIFICATION OF STANDARDIZATION PROCEDURES

Standardization refers to transformations that involve an adjustment of means and/or standard deviations of either individuals or groups. Therefore, these procedures can be ordered depending on whether they used only the means, only the standard deviations, or both (see first column in Table 1). Let us first turn to adjustments using means. These transformations are usually in the form of deviation scores whereby the mean across variables or individuals will be subtracted from some raw mean of either a variable or individual. Additionally, it is possible to adjust the resulting deviation score by measures of dispersion, for example, standard deviations. Standardization can involve adjustment of means of either individuals, groups, or both using either the mean across variables for each individual or

TABLE 1

Classification of Possible Standardization Procedures

	Within-Subject (Adjustment Across Variables for Each Individual)	*Within-Group (Adjustment Across Individuals for One Variable)*	*Within-Culture (Adjustment Across Individuals and Variables)*	*Double*
Adjustment using means	$y' = x - mean_{individual}$ Ipsatization	$y' = x - mean_{group}$ Group mean centering	$y' = x - mean_{culture}$ Grand mean centering	$y' = x - mean_{individual}$ $y'' = y' - mean_{y'culture}$
Adjustment using dispersion indices (commonly standard deviation)	$y' = x/dispersion_{individual}$	$y' = x/dispersion_{group}$	$y' = x/dispersion_{culture}$	$y' = x/dispersion_{individual}$ $y'' = y'/dispersion_{y'culture}$
Adjustment using means and dispersion indices	$y' = (x - mean_{individual})/dispersion_{individual}$ Ipsatisation	$y' = (x - mean_{group})/dispersion_{group}$ z-transformation	$y' = (x - mean_{culture})/dispersion_{culture}$	$y' = (x - mean_{group})/dispersion_{group}$ $y'' = (x - mean_{y'group})/dispersion_{y'group}$
Adjustment using covariates	Partial correlation: $r_{12.C} = (r_{12} - r_{1C}r_{2C})/\surd[(1 - r^2_{1C})(1 - r^2_{2C})]$ C = $mean_{individual}$	Partial correlation: $r_{12.C} = (r_{12} - r_{1C}r_{2C})/\surd[(1 - r^2_{1C})(1 - r^2_{2C})]$ C = $mean_{group}$	Partial correlation: $r_{12.C} = (r_{12} - r_{1C}r_{2C})/\surd[(1 - r^2_{1C})(1 - r^2_{2C})]$ C = $mean_{culture}$	
	One-way analysis of covariance (based on Everitt & Wykes, 1999): $y_{ij} = \mu + a_i + \beta x_{ij} + e_{ij}$	One-way analysis of covariance (based on Everitt & Wykes, 1999): $y_{ij} = \mu + a_i + \beta x_{ij} + e_{ij}$	One-way analysis of covariance (based on Everitt & Wykes, 1999): $y_{ij} = \mu + a_i + \beta x_{ij} + e_{ij}$	
	y_{ij} = value of the dependent variable for the jth individual in the ith group μ = overall mean a_i = group effect x_{ij} = value of the covariate ($mean_{individual}$) for the individual β = regression coefficient of the dependent variable on the covariate e_{ij} = residual or error term	y_{ij} = value of the dependent variable for the jth individual in the ith group μ = overall mean a_i = group effect x_{ij} = value of the covariate ($mean_{group}$) for the individual β = regression coefficient of the dependent variable on the covariate e_{ij} = residual or error term	y_{ij} = value of the dependent variable for the jth individual in the ith group μ = overall mean a_i = group effect x_{ij} = value of the covariate ($mean_{culture}$) for the individual β = regression coefficient of the dependent variable on the covariate e_{ij} = residual or error term	

across individuals within a group, or both. Therefore, standardization procedures can also be ordered depending on what the source of the information is (e.g., individuals, groups, culture; see first row of Table 1). In combination with the type of information used (means, standard deviations, or both), there are a number of possibilities for adjusting raw scores. In the following, I will review the different possibilities.

First, within-subject standardization refers to the adjustments of scores for each individual using the mean for that individual across all variables (Hofstede, 1980). The average across a subset or all variables for that particular individual is subtracted from each individual's raw score. Therefore, the resulting score is the relative endorsement of this item or the relative position of the individual on a variable in relation to the other scores (Hicks, 1970). The mean across variables for this individual will average to zero. This procedure is also called ipsatization (Hicks, 1970). These deviation scores might be further adjusted for differences in the variation of the answers around the mean by dividing the resulting score by the standard deviation across variables for that individual. Schwartz (1994) proposes a related strategy at a culture level:

> The importance ratings for all the single values included in a type are averaged. These scores were corrected for sample differences in scale use by subtracting from each sample's scores the average rating that the sample gave to all 56 values. The scores were then rescaled to reflect the original –1 to +7 importance scale by adding to them the average rating given to all 56 values across all samples. (p. 100)

Consequently, he treated countries as if they were individuals and standardized across individual variables (values) for each country and then used a rescaling procedure to retain the original scale.

Second, answers can be adjusted using the group mean. There are several variants. The most common form is the usual computation of *z* scores whereby the group mean is subtracted from the variable raw score and then divided by the standard deviation (Howell, 1997). Therefore, the resulting score is the relative endorsement or position of one specific individual on one variable relative to the endorsement or position of other individuals in that group. The mean across individuals is zero and, assuming a normal distribution of responses, the resulting standard deviation will be 1. An adjustment procedure using only the mean across individuals is also possible and is commonly referred to as centering (Aiken & West, 1991).

A related strategy is within-culture standardization (Bond, 1988; Leung & Bond, 1989). However, instead of using the group mean across one item (as done in *z* transformation), the mean across all items and all individuals (the grand mean) is used. In the method literature, this is sometimes referred to as grand mean centering (Raudenbush & Bryk, 2002). Again, it is possible to adjust this deviation score by dividing it by the standard deviation across items and individuals (the grand standard deviation).

Leung and Bond (1989) introduced a combination of within-subject and within-culture standardization, which they labeled double standardization. First, scores are adjusted within the individual (within-subject standardization), and then the resulting scores are adjusted within the group (within-culture standardization). Therefore, the means for each individual across variables and the mean for each variable across individuals will be zero. Assuming normality of the raw data, the adjustment using the standard deviation should yield standard deviations of 1 for both individuals across variables and variables across individuals.

It is important to note that these standard procedures can be done at an item and construct level. One may standardize individual items using the relevant mean across either items or individuals or one may standardize the mean for one specific construct using the relevant mean across either items, constructs, or individuals. As an example, focusing on the Schwartz Value Survey (Schwartz, 1992), one may standardize each individual item of the 56 values inventory within each individual using the mean across the 56 items for that individual. Alternatively, one may use the mean score for 1 of the 10 value types (e.g., hedonism) and adjust it using the overall mean across all values or value types for that individual. The same procedure might be used for adjustment within groups at either the item or construct or scale level.

The previous standardization procedures adjusted for mean differences, and an adjustment for dispersion or extreme rating was optional. It is also possible to adjust solely for differences stemming from extreme rating, which are likely to result in greater dispersion indices (assuming equal numbers of positively and negatively phrased items; see Chun, Campbell, & Yoo, 1974). Using the differentiation between within-subject and within-group or culture again, the raw score for an individual can be adjusted by dividing his or her score for one specific variable by the standard deviation across a set of or all variables for that individual. This would be equivalent to within-subject standardization using standard deviations only. Alternatively, within-group or culture standardization would involve an adjustment of means by the standard deviation for the group. Assuming a normal distribution of raw scores, both procedures should result in standard deviations of 1 (across individuals or groups, respectively). Kashima, Siegal, Tanaka, and Kashima (1992) proposed a related strategy to account for dispersion differences:

> To make dispersions constant across cultures when controlling for extreme response sets, the difference between a raw rating score and the neutral point of the scale (3 for five-point and 4 for seven-point scales) was divided by the standard deviation for the scale. (p. 117)

They argue that this procedure equates the extent to which participants departed from the midpoint of the scale. They proposed it as a within-subject procedure, but similarly the relevant standard deviation could be the group standard deviation rather than the standard deviation for the individual. Adjustments for measures of dispersion generally tend to use standard deviations; however, it is also possible to use the range or the minimum and maximum for adjustment procedures (Kurman & Sriram, 1997).

A final class of procedures has been proposed based on the covariance analysis (see the final row of Table 1). Although this method is not new (Russell, 1979), it has been made popular in the cross-cultural literature by Schwartz (1992). Different from his treatment of the data at a culture level, he recommends an alternative strategy based on partial correlation and analysis of covariance at an individual level of analysis:

> We propose to use each individual's mean importance rating for the 56 core values as a covariate in comparisons of group means, or as a third variable whose effect on the correlations between value priorities and other variables is controlled through partial correlation. (p. 53)

This procedure has also been used for factor analysis (Hui & Yee, 1994). Conceptually, this strategy is similar to within-subject standardization because the dependent variable of interest is regressed first onto the overall mean across items for an individual, therefore accounting for systematic differences in scale use for each individual. Although Schwartz recommended using the mean for each individual across variables, it is theoretically also possible to use either the mean across individuals within a group (similar to *z*-transformation, centering) or the mean across variables and individuals (similar to within-culture adjustment).

Therefore, there are four major groups of standardization procedures. These four groups are (a) adjustment of means, (b) adjustment for dispersion (e.g., using standard deviation), (c) adjustments using means and dispersions, and (d) covariate analysis. These procedures can be further divided by identifying which statistical information they use for standardization. As outlined in the previous sections, data can be standardized within-subject, within-group, and within-culture. The first uses means or standard deviations across variables for each individual, the second uses means or standard deviations across individuals for each variable, and the last uses means and standard deviations across both variables and individuals within a sample coming from the same cultural background. Double standardization uses within-subject and within-culture information for adjustment. Table 1 displays the resulting table together with the appropriate formulas.

It should be noted that most studies have used the mean or standard deviation across all items or used a subset of items based on their theoretical and conceptual interest (e.g., all individualism-collectivism items) when adjusting for response bias. Therefore, it is assumed that response tendencies are consistent across items and constructs. This strategy represents a rather undifferentiated approach to score adjustment. Various researchers have developed more sophisticated methods for detecting response patterns. Using equal numbers of positively and negatively phrased items from conceptually related and unrelated scales, more sophisticated and precise adjustment strategies become possible. Watkins and Cheung (1995) describe various possibilities for developing control scores. For example, comparing answers to positively and negatively phrased items from the same construct, it is possible to evaluate whether individuals responded more positively or negatively, irrespective of the item content. Such an index can then be used for adjustment in subsequent analyses (e.g., Crittenden, Fugita, Bae, Lamug, & Lin, 1992; Hofstee, Ten Berge, & Hendriks, 1998).

The rationale for using either means or standard deviations when adjusting raw data is also different. Adjustment using means is typically indicated if researchers expect acquiescent bias. Therefore, the mean across all items for one cultural group is consistently higher or lower compared with the means from another group, and adjusting for these mean differences might be advocated (Hofstede, 1980). The rationale for using standard deviations is to adjust for extreme response bias (Kashima et al., 1992). The most straightforward scenario includes a relative balance of positively and negatively phrased items (Chun et al., 1974). Observing an overall difference in standard deviations would make extreme response tendencies a possible explanation. Using only items phrased in one direction (either only positively or only negatively) makes detection of extreme responding more difficult, especially considering that both types of response tendencies are often found together (Smith, 2004). In such a case, adjustment using both means and standard deviations might be appropriate.

REVIEW OF STUDIES PUBLISHED IN *JCCP*

In the following sections, I will review studies published in *JCCP* between 1970 and 2002 to evaluate how often authors have used these standardization procedures. I will also examine which statistical techniques were subsequently employed by these authors and what reasons researchers cited for standardization.

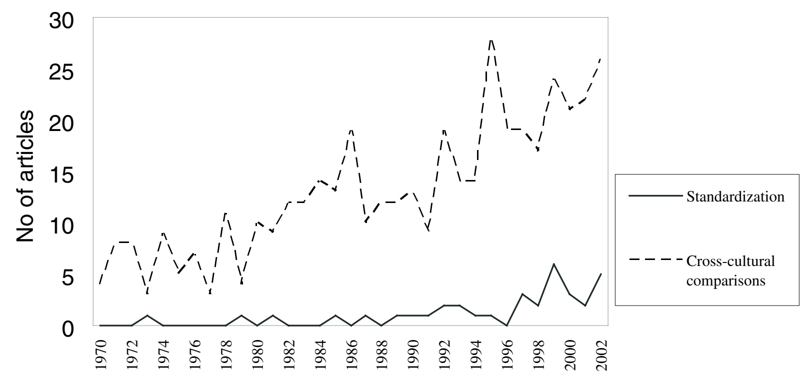

Figure 1: Frequency of Standardization in *JCCP* Between 1970 and 2002

All method and results sections of articles published during this period were read to determine whether authors used some form of standardization when analyzing data based on rating scales in their analysis. All together, there were 35 articles that used some form of standardization procedures. Figure 1 displays the number of cross-cultural comparisons using some form of questionnaire measure as well as the number of articles using standardization procedures. There were only three studies prior to 1985, but since then there has been a relatively steady increase in the use of standardization procedures. Another point worth noting is that the number of studies comparing ratings from two or more cultural or ethnic groups has steadily increased over the years. However, the increase of the use of standardization since the mid-1980s is substantial. I calculated the percentage of studies within each year that used standardization procedures. I then correlated this percentage (using data since 1985) with the publication year. The resulting rank-order correlation was .58, which is significant at a .05 level. Therefore, there is a trend to use standardization procedures more often in more recent studies.

Examining the type of standardization analysis used, nine articles employed within-subject standardization using means and standard deviations; seven articles used double standardized scores, and five articles conducted within-subject standardization using participants' means only. Three studies employed within-culture standardization using only culture means, whereas five studies used both culture means and standard deviations. Two studies computed *z*-scores, three studies used some form of covariate analysis, and one study used a variant of the dispersion adjustment (using the standard deviation of individuals). Finally, one study used a different standardization technique, and three studies could not be classified due to missing or unclear information. It should be noted that it was often quite ambiguous whether researchers used either only the mean or both means and standard deviation for standardization. See Table 2 for the list of studies.

Next, it was investigated which statistical techniques were used by the authors. Here, I will focus only on the three standardization procedures that were used most often (within-subject using means and both means and standard deviations, double standardization).

TABLE 2

List of Articles Using Standardization (Ordered by Publication Year)

	Standardization Procedure	*Data Analysis*	*Sample*
Adams & Osgood (1973)	Within-subject, means, and standard deviations	Ranking, distances matrix	20 countries
Munroe (1979)	Within-subject, means, and standard deviations	Factor analysis using standardized and unstandardized scores; ANOVA with standardized scores	Racial groups in Zambia, Zimbabwe
Fioravanti, Gough, & Frere (1981)	Within-culture, means, and standard deviations	ANOVA	United Kingdom, France, Italy
Bond, Wan, Leung, & Giacalone (1985)	Centering	Factor analysis, ANOVA (but unclear whether standardized scores)	United States, China
Chinese Culture Connection (1987)	Within-subject, means, and standard deviations	Factor analysis, correlations	22 countries
Boehnke, Silbereisen, Eisenberg, Reykoowski, & Palmonari (1989)	Within-subject, means, and standard deviations	MANOVA	West Germany, Italy, United States, Poland
Wagner, Kirchler, Clack, Tekarslan, & Verma (1990)	Within-subject, means, and standard deviations	ANOVA	Austria, India, Turkey, United States
Minsel, Becker, & Korchin (1991)	Within-subject, means, and standard deviations	Factor analysis, ANOVA	France, Germany, Greece, United States
Feather, Volkmer, & McKee (1992)	Within-subject, means, and standard deviations	ANCOVA, rank order correlations; discriminant function analysis	Australians, Australian Baha'is, and expatriate Iranian Baha'is
Crittenden, Fugita, Bae, Lamug, & Lin (1992)	Covariate analysis	ANCOVA	Korea, Philippines, Taiwan, United States
Phalet & Claeys (1993)	Double standardization	Factor analysis, ANOVA based on factor scores, multidimensional scaling, correlations	Turkey, Belgium
Triandis et al. (1993)	Double standardization	Factor analysis, discriminant function analysis	10 countries
Fogarty & White (1994)	Within-subject, means, and standard deviations	MANOVA, ANOVA	Australian Aboriginal and non-Aboriginal students
Yamaguchi, Kuhlman, & Sugimori (1995)	Double standardization, Kashima et al. (1992) method (dispersion procedure)	Factor analysis with double standardized scores, ANOVA with Kashima et al. procedure	United States, Japan, Korea
Kurman & Sriram (1997)	Dispersion adjustment for possible range of answers	ANOVA	Israel, Singapore

Marshall (1997)	z-transformation	ANOVA	Indonesia, New Zealand
Matsumoto, Weissman, Preston, Brown, & Kupperbusch (1997)	Within-culture, means, and standard deviations[a]	Comparison of ANOVA results with standardized and unstandardized data	United States, Japan, South Korea, Russia
Schultz & Zelezny (1998)	Covariate analysis	Partial correlation	Mexico, Nicaragua, Peru, Spain, United States
Hasegawa & Gudykunst (1998)	Possibly double standardization	Factor analysis, ANOVA with raw scores	Japan, United States
Ohbuchi, Fukushima, & Tedeschi (1999)	Within-subject, means, and standard deviations	ANOVA, regression with raw scores	Japan, United States
Goodwin, Nisharadze, Kosa, & Emelyanova (1999)	Possibly double standardization	Factor analysis	Russia, Hungary, Georgia
Singelis, Bond, Sharkey, & Lai (1999)	Double standardization	Factor analysis	United States, Hawaii, Hong Kong
Grimm, Church, Katigbak, & Reyes (1999)	Within-subject, means	Factor analysis, Mann-Whitney U test	Philippines, United States
Aycan, Kanungo, & Sinha (1999)	Double standardization	ANOVA, not clear whether standardized scores were used for regression analysis	India, Canada
Stewart, Bond, Deeds, & Chung (1999)	Covariate analysis	Partial correlation, ANCOVA	Asian and Caucasian teenagers and mothers in Hong Kong
Merrit (2000)	Within-subject, possibly only mean	correlation	19 countries
Schaffer, Crepaz, & Sun (2000)	z-transformation	ANOVA	United States, Taiwan
Landis, & O'Shea (2000)	Within-culture, possibly means only	3-Mode Factor Analysis	Canada, Israel, United States, United Kingdom, Denmark
Kirkman & Shapiro (2001)	Within-culture, possibly means only	Regression analysis	United States, France, Belgium, Philippines
Ralston et al. (2001)	Within-culture, means, and standard deviations[a]	MANCOVA	Germany, Hong Kong, India, Mexico, the Netherlands, United States
Nelson & Shavitt (2002)	Within-subject, possibly means only	MANOVA, correlation	Denmark, United States
Van de Vliert & Janssen (2002)	Within-culture, means, and standard deviations[a]	Multidimensional scaling, regression	42 countries
Seki, Matsumoto, & Imahori (2002)	Within-culture, means, and standard deviations	MANOVA	Japan, United States
Smith, Peterson, & Schwartz (2002)	Within-subject, means	Factor analysis, correlation	47 countries
Harrington & Liu (2002)	Within-subject standardization, means	U test, correlation	Maori, White New Zealander

a. Information obtained from authors.

Studies using these three procedures mostly conducted some form of ANOVA or MANOVA (11 studies) and factor analysis (12 studies). Five studies computed correlations, 4 studies reported nonparametric tests, and 2 studies used discriminant function analysis. Please note that studies were coded more than once if they employed more than one technique.

Van de Vijver and Leung (1997) argued that researchers should provide a rationale or theoretical justification why it is necessary to employ justification. Nearly all researchers (24 studies) referred to cross-cultural differences in scale use or response bias when talking about standardization. One of the most often cited studies was Bond (1988). Bond (1988) did not refer to any specific studies that found cross-cultural differences in response bias rather than noting that such biases can exist. Some authors (e.g., Aycan, Kanungo, & Sinha, 1999; Minsel, Becker, & Korchin, 1991) noted that overall differences might also be caused by substantial factors related to cultural variables rather than reflecting bias, but nevertheless continued to standardize their data prior to analysis.

Only few researchers actually discussed why a higher mean indicates bias rather than some meaningful theoretical difference. Schwartz's (1992) research on human values is a good example. Based on his theory, he pointed out that the overall average score should be similar across individuals and cultures because his instrument is intended to represent a comprehensive set of human values (Van de Vijver & Leung, 1997). Overall mean differences, therefore, represent some bias rather than meaningful variation. This argument was also used by Smith, Peterson, and Schwartz (2002) in their study of event management across cultures. Another noteworthy example is the study by Harrington and Liu (2002). They argued that standardized scores create "a limited view of the self, whereby each person has the same fixed amount of self-construal resources to allocate, whereas raw scores allow some persons or groups to have higher self-ratings across the domains than others" (p. 47). Therefore, they theoretically assumed a limited resource perspective of the self, which indicates that overall differences across samples are an indicator of bias. Researchers are urged to provide a theoretical rationale why mean differences across constructs are an indicator of bias rather than some meaningful variation. This is especially important in light of those recent studies showing that mean differences might be due to culturally determined communication patterns rather than bias (Smith, 2004; Van Hemert et al., 2002).

Four researchers used standardization to obtain a culture-free or etic correlation matrix to be used in factor analysis (see below). Three studies used standardization after they observed consistent mean differences irrespective of variable content. Two further studies did not provide any rationale, and one study attempted to provide a comparable metric for scales with different response format. Therefore, only few researchers provided a theoretical rationale for their standardization.

In summary, there has been a steady increase in the use of standardization procedures since 1985. The most common standardization procedures are within-subject standardization and double standardization. Most researchers used these standardized scores to conduct some form of parametric analysis, mostly ANOVA or factor analysis. Finally, the main cited reason for standardization was to eliminate or reduce response bias, although researchers did not theoretically discuss why such differences are aspects of bias rather than substantial and meaningful variation. A smaller number of studies used standardization to derive a culture-free or etic correlation matrix to be used in factor analysis. With these findings in mind, I will turn next to statistical properties and implications for data analysis.

STATISTICAL PROPERTIES OF STANDARDIZATION PROCEDURES

In the following section, I will review statistical properties of within-subject standardization (including double standardization), within-group standardization, and covariate analysis, and will highlight some implications for both correlational and analysis of variance designs. Within-subject standardization and double standardization are the most commonly used standardization procedures; therefore, I will dedicate more space to these procedures.

IMPLICATIONS OF WITHIN-SUBJECT STANDARDIZATION AND DOUBLE STANDARDIZATION FOR CORRELATIONS AND FACTOR ANALYSIS

As noted earlier, within-subject standardization yields a mean of 0 for each individual and a standard deviation of 1 (if divided by the standard deviation of the individual and assuming normality of the measures). These properties characterize ipsative scores. Hicks (1970) was among the first to point out their limitations for correlational techniques. Since then, a number of theoretical and empirical papers have further developed these arguments (e.g., see Baron, 1996; Bartram, 1996; Chan, 2003; Closs, 1996; Cornwell & Dunlop, 1994; Johnson, Wood, & Blinkhorn, 1988; Tenopyr, 1988). I will only provide a brief review of these summaries related to correlations and factor analysis, and will then add some more concerns relating to ANOVA and MANOVA designs in the next section.

The sum of ipsative scores is a constant for each individual (constant-sum constraint) and "each score for an individual is dependent on his own scores on other variables, but is independent of, and not comparable with, the scores of other individuals" (Hicks, 1970, p. 167). Implications of these properties are that the sum of variances and covariances is zero in every row and column of the covariance matrix. This results in a singular matrix with no regular inverse which creates problems for factor analysis (Chan, 2003). Furthermore, at least one of the $k - 1$ (with k being the number of variables) covariance terms in each row and column is forced to be negative, irrespective of substantive relationships. Therefore, at least one covariance (or correlation) is not determined by the true relationship between constructs, but rather methodological artifacts. The average item-intercorrelation can be predicted by the number of variables standardized (average $r = -1/[k - 1]$, where k is the number of variables, assuming equal variances; Hicks, 1970). Finally, the sum of covariances and correlations (if the ipsative variances are equal) with some external variable will average to zero across all scales on which the standardization was based (Chan, 2003; Hicks, 1970). This has serious implications for any statistical analysis based on correlations.

For example, reliabilities are sometimes seriously inflated (Tenopyr, 1988) or deflated (Bartram, 1996). Bartram (1996) used hypothetical data to show that these problems are aggravated when the number of scales is less than 10 and the correlation between the original scales is greater than .30. Cornwell and Dunlop (1994) showed that correlation matrices based on ipsatized scores are not suitable for factor-analytical techniques. They will yield spurious bipolar factors. Closs (1996) also argued strongly against using ipsatized scores in factor analysis. He compared intercorrelation matrices based on the same individuals, but either in ipsative or nonipsative form. The intercorrelations of the ipsative scores were mainly negative (see Hicks, 1970), whereas the nonipsative scores were positively correlated. Consequently, ipsatized scores are not suitable for factor analytical techniques, the reliabilities are generally affected, and correlations (and regressions) with external variables have to be interpreted with great caution.

In light of the previous review, the common argument that correlation matrices based on double standardized scores yield so-called culture-free (Hasegawa & Gudykunst, 1998; Leung & Bond, 1989; Singelis, Bond, Sharkey, & Lai, 1999; Triandis et al., 1993) or etic dimensions (Yamaguchi, Kuhlman, & Sugimori, 1995) has to be seriously questioned. The previous points are especially important in relation to the common procedure of standardizing variables that are highly correlated with each. Hasegawa and Gudykunst (1998) standardized items relating only to silence, whereas Nelson and Shavitt (2002), Yamaguchi et al. (1995), and Triandis et al. (1993) standardized responses based on only individualism-collectivism items. Assuming that these variables are theoretically related, at least some of the observed variation in the standardized data is due to purely methodological artifacts rather than substantial relationships.

To demonstrate this point, I administered an 18-item version of the horizontal-vertical individual-collectivism scale (Triandis & Gelfand, 1998) to 109 second-year psychology students as part of an in-class exercise. The modal age of the participants was 19; about 25% were male and 76% were born in New Zealand. I tested whether I could recover the initial factor structure in the same data set after within-subject standardization. The average item intercorrelation was .11, with a maximum of .66 and a minimum of –.27. According to Bartram's (1996) analysis, this data set might be suitable for standardization. Six factors with eigenvalues greater than 1 were found in the standardized and the unstandardized data matrix. The scree test indicated the presence of four factors in the raw data matrix. These four factors derived from the raw score correlation matrix represented horizontal-individualism, vertical-individualism, vertical-collectivism, and horizontal-collectivism dimensions. Only the item "It is important to me that I respect the decisions made by my groups" loaded on both collectivism dimensions. Consequently, the structure can be found in the raw data. Therefore, I extracted six factors and four factors and rotated the structure of the standardized matrix to the unstandardized matrix. I compared the solution using Tucker's Phi. Values of .95 or higher are normally seen as indicators of good factorial agreement (Van de Vijver & Poortinga, 2002). The values for the six-factor solution were .75, .96, .87, .73, .74, and .94. Therefore, except for the second factor, the factors were not adequately recovered in the standardized matrix.

The four-factor solution fared somewhat better. The values of Tucker's Phi were .99, .99, .79, and .42. The first two factors were adequately recovered, but the agreement of the other two factors was not acceptable. These two factors correspond to the vertical-collectivism and the horizontal-individualism dimension. Therefore, two of the four extracted factors were not adequately recovered. So-called culture-free correlation matrices based on within-subject and double standardization yield different factor structures even within the same data set, and interpreting resulting factor structures is highly ambiguous (Closs, 1996; Cornwell & Dunlop, 1994).

Chan (2003) proposed a method based on confirmatory factor analysis which allows the use of ipsative data matrices. The procedure set forth by Chan involves a number of constraints to be placed on the factor loading and error covariance matrix. Chan reports an adequate recovery of the initial factor structure; however, it should be noted that the fit indices differ considerably between the raw and standardized matrix solution. Further research is needed to investigate this promising possibility.

Therefore, research so far suggests that interpreting results from correlational analyses, such as factor analysis and regression that is based on ipsative scores, is ambiguous. This problem is aggravated if the scales used for ipsatization are highly intercorrelated. In these cases, ipsatization is likely to produce spurious results that in most cases do not correspond to substantial relationships (see Closs, 1996; Cornwell & Dunlop, 1994). If researchers are interested in investigating the underlying structure of ipsative data matrices, nonparametric techniques such as multidimensional scaling could be used. Multidimensional scaling (Kruskal & Wish, 1978), using distances, is an appropriate alternative for analyzing ipsative data. Furthermore, multidimensional scaling can be used to assess structural relations among variables instead of factor analysis if researchers are concerned that response biases might obscure structural relationships in the data, because this technique is not influenced by overall score level differences in different groups.

IMPLICATIONS OF WITHIN-SUBJECT STANDARDIZATION AND DOUBLE STANDARDIZATION FOR ANALYSIS OF VARIANCE DESIGNS

What are implications for analysis of variance designs? Some researchers found different results when using standardized data compared with unstandardized data (e.g., Matsumoto, Weissman, Preston, Brown, & Kupperbusch, 1997), whereas others (e.g., Kirkman & Shapiro, 2001) found identical patterns with standardized and unstandardized data. To highlight one potential reason for this difference, let us consider the example of a mouse and an elephant. Assume someone measured the extremities of both animals and used within-subject (within-animal) standardization. If the researcher would now proceed to compare the length of, let us say, the legs, probably no significant differences would be found. This is despite the fact that the legs of an elephant and a mouse are surely different. This is because all the measures are related to the size of the whole animal. Therefore, the first implication is that if dimensions are interrelated or have a common underlying cause (e.g., overall body size), comparison of standardized measures in absolute terms (which leg is longer, the leg of the elephant or the leg of the mouse?) will give misleading results. This is related to findings by Bartram (1996) in the context of factor analysis, where he found that standardization might be applicable in situations where many dimensions are measured that are not strongly intercorrelated. Whereas Matsumoto et al. (1997) used presumably highly intercorrelated social values related to individualism-collectivism, Kirkman and Shapiro (2001) used four different value dimensions as well as variables at a team level that are less likely to be highly correlated. As I discussed in the section on factor analysis, ipsatization will force some variables to take on lower or higher values than in the raw data, and if the original items are highly correlated, this might produce spurious results.

A second implication is that if we compare the tail of the mouse and the elephant using ipsative measures, we would probably conclude that the mouse's tail is significantly longer than the tail of the elephant. It is important to note that this comparison makes sense only if we consider the length of the tail relative to the overall size of both animals. Obviously, relative to the overall size of the mouse and the elephant, the mouse's tail is longer than the tail of the elephant.

This analogy nicely illustrates characteristics of ipsatization that have been noticed by various authors. Ipsative scores are not comparable across individuals but might only be used for intraindividual comparisons. For example, the seminal article by Hicks (1970) concluded that "in summary, one may state that scores originally obtained as ipsative measures may be legitimately employed only for purposes of intraindividual comparisons" (p. 168). Chan (2003) concurs:

> As with ipsative measurement, there is a scale for every individual within which variations of behaviour occur. They are therefore inappropriate for inter-individual comparisons as two per-

> sons having an identical set of ipsative responses could be very different in terms of their overall amount of attitude or behavior. In fact, ipsative scores can only reflect the intra-individual (relative) differences across the variables and they are only appropriate for intra-individual comparisons. (p. 100)

If researchers are interested in the relative importance of variables within samples (which is basically what is achieved with ipsatization), then repeated measures ANOVA or mixed-model ANOVAs using the raw data might be more appropriate, and the resulting effect sizes of the within-culture comparison might be compared across samples (for an example, see Gudykunst et al., 1992). Ipsative scores could also be compared using appropriate nonparametric tests (for an example, see Grimm, Church, Katigbak, & Reyes, 1999). Harrington and Liu (2002) also used nonparametric tests. Their study is also noteworthy because, as discussed previously, they provided a theoretical rationale for the meaning of the standardized scores. Observed mean differences were interpretable because the authors assumed that the self has a limited size and some aspects are relatively more important than others within each person's self-concept. Finally, if only the order of variables is of interest, the rank order might be compared across cultural samples (see Church & Katigbak, 1992).

A final note related to the use of MANOVA and discriminant analysis. Both techniques involve highly sophisticated and complex matrix computations and rely on inverse matrices (see Field, 2000). As noted above, ipsative data matrices have no regular inverse, which is likely to create problems for MANOVA and discriminant function analysis. It should also be noted that Triandis et al. (1993) were unable to replicate discriminant functions derived from unstandardized scores with ipsatized scores using the same data sets across 10 cultural samples. Considering the complex statistics and the problems outlined above, authors might be advised to refrain from using multivariate techniques with ipsative scores pending further investigation of the impact of ipsatization on these techniques.

IMPLICATIONS OF WITHIN-GROUP AND WITHIN-CULTURE STANDARDIZATION

These standardization procedures are less commonly used and do not share the same statistical properties as ipsative scores yielded through within-subject standardization. Centering around the group mean or grand mean is a common procedure for various statistical techniques such as moderated multiple regression (Aiken & West, 1991) or hierarchical linear modeling (Raudenbush & Bryk, 2002). It can also be used for exploratory factor analysis to remove patterning effects (e.g., Leung & Bond, 1989). Overall, these standardization methods can be used; however, the interpretation of the resulting scores is crucial. Depending on the standardization used (within-group, grand mean) and the number of variables and their intercorrelation (for within-culture standardization), the interpretation of results changes. The classical example is *z*-transformation. Variables on different metrics can be compared in relation to the relative standing of individuals or groups compared to the mean. However, separate *z*-transformations within each sample will basically eliminate any mean differences across samples. The example of the mouse and elephant might be relevant again. The resulting comparison of means has to take into account which variables were used for standardization and what the substantive relationship between these variables is.

In the context of regression, mean centering is used to provide a meaningful metric for moderation analysis (see Aiken & West, 1991). In hierarchical linear modeling (Raudenbush & Bryk, 2002), a method conceptually similar to regressions but using variables at an individual and group (or culture) level, group-mean and grand-mean centering are routinely employed to distinguish individual-level and group (culture) level effects depending on the expected variation of effects at an individual and group (cultural level). However, the interpretation of the results changes depending on the centering used. I will use the link between self-esteem and life satisfaction as a somewhat simplified example. I assume self-esteem is positively correlated with life satisfaction, but different groups have different life satisfaction levels.[1] If researchers are interested in the effect of self-esteem on life satisfaction of individuals irrespective of the group differences, then within-group centering could be used to show the relationship at an individual level while controlling for the group level differences. The simple regression weight would show this individual level effect. However, if contextual effects are of interest, for example, the influence of group level self-esteem on the life satisfaction of individuals, controlling for self-esteem at an individual level, researchers may use grand-mean centering. The resulting regression weight in hierarchical linear modeling would represent the effect of group differences on satisfaction levels of individuals. Therefore, standardization can be necessary and useful for researchers, and the importance lies in the appropriate interpretation of the results obtained (see Raudenbush & Bryk, 2002, pp. 134-149).

IMPLICATIONS OF COVARIATE ANALYSIS

Schwartz (1992) recommended partialing the mean across the 56 values in his value survey as a way of controlling for response bias. As discussed before, he theoretically supported this recommendation by pointing to his theoretical structure, which would indicate that mean differences across values are bias rather than substantial variation (Van de Vijver & Leung, 1997). Although this is an appealing strategy, it may not be without limitations in different data sets and with different instruments. As can be seen in Table 1, the rationale underlying all types of covariance analyses is to use residuals of regressions purified of the effect of a nuisance variable (response bias). One way is to think of a simple regression of the single item on the composite scale (including the item). All the variance in the single item that is not accounted for by the composite scale (unexplained variance is the deviation from the predicted regression line or the residual of the regression) is then used in the following analysis (correlation or analysis of variance). But what happens if an item is regressed against itself (or against itself together with other items)? In an extreme case singularity will be encountered, meaning that there is no residual because the dependent variable is perfectly predictable from the predictor variable. Consequently, this kind of analysis takes out variance which is not caused by response bias but the true endorsement of the variable. One remedy would be to eliminate the respective item (or sets of items) from the partial correlation (Russell, 1979). As an example, if you are interested in the relation between hedonism and stimulation, you could partial the mean of all variables without the hedonism and stimulation values. Nevertheless, this procedure is likely to result in the problems as outlined above, if other value types (e.g., adjacent value types such as power or self-direction) are highly correlated with hedonism and stimulation and true variance that is shared by these value types is removed. This also might create the problem of multicollinearity, resulting in unstable regression weights.

Another problem is the number of items. Instruments often consist of different numbers of items. Let us assume we have 10 items, 8 items measuring one construct and the remaining 2 items form a second, independent construct. If the overall endorsement is taken as a response bias measure, the answer of the 8 items is adjusted by using the answers of 2 additional items, whereas the response of the 2 items will be adjusted for the answers of the other 8 items. To overcome this problem, the instruments should be of comparable length.

Given the large number of values in the Schwartz value survey, the circumplex structure resulting in positive, negative, and zero correlations among values and the apparent comprehensiveness of the instrument cross-culturally, partialing the mean might be appropriate. However, this technique might yield biased results when smaller numbers of interrelated constructs (sharing substantial variance) are used and the number of items differs substantially across constructs.

SUMMARY AND DISCUSSION

Dealing with cross-cultural response patterns is arguably one of the most challenging issues in cross-cultural survey research. Standardization has been proposed as a mean to overcome this problem; however, there has been confusion about the meaning of this term. Therefore, this article provides a general classification scheme for standardization procedures for eliminating and reducing response bias in cross-cultural research. It gives an overview of the various standardization possibilities and helps to clarify the previous inconsistent and ambiguous use of the term *standardization* in the cross-cultural literature. Using this classification scheme, a review of empirical research in *JCCP* has shown that authors increasingly rely on standardization. Within-subject and double standardization procedures are the most common procedures. Despite citing eliminating potential bias as a reason for standardization, most researchers do not discuss why observed mean differences constitute bias rather than substantial variation that might be linked to their topic of interest. This is a major shortcoming of previous research, and researchers should consider alternative meanings of (unwanted) cross-cultural differences in more depth.

Furthermore, within-subject and double standardization procedures create ipsative scores that have various undesirable properties, especially for statistical techniques involving correlations. The author would caution researchers against the use of within-subject and double standardized data matrices for factor analysis, especially if constructs in the raw data matrix are positively correlated. So-called culture-free or etic factor analyses are likely to produce spurious method factors (Cornwell & Dunlop, 1994). Researchers interested in the underlying structure of their standardized data matrix might use multidimensional scaling instead. Researchers interested in correlation are advised to consider rank-order correlations instead of parametric Pearson correlation with standardized data. They should also note that the sum of correlations between ipsative measures and external unstandardized criteria will be close to zero. This methodological constraint might alter interpretations of the results and researchers should consider the theoretical implications. Finally, more research is needed to understand the performance of ipsative data when using other multivariate statistics such as MANOVA, discriminant function analysis, and structural equation modeling.

Comparison of means based on ipsative scores are difficult to interpret, and the exact interpretation depends on a number of factors (such as the number of variables used for standardization and the interrelation between these variables). Researchers are encouraged to consider theoretical implications of ipsatization because this introduces a constant-sum constraint. Such constraints can make sense if a theory assumes an upper limit (as in the case of self; see Harrington & Liu, 2002) or the theoretical interest is in the relative importance of scores in relation to other variables. Using standardized data, nonparametric tests should be conducted (e.g., Grimm et al., 1999). Furthermore, other options are available if there is concern with cross-cultural response biases and researchers are only interested in relative rather than absolute differences. Both within-subject and mixed-model designs might be used to overcome some of these problems (e.g., Gudykunst et al., 1992). Using within-subject and mixed-model designs, the overall response tendency for each individual is taken into account, therefore correcting for response biases exhibited by individuals. For example, if individuals are interested in differences between individualism and collectivism in two different cultures, they might use separate scales for individualism and collectivism and then use the reported importance of individualism compared with collectivism as within-subject factor and culture as a between-subject factor. The interaction between the within-subject and between-subject factor would indicate whether the means for individualism and collectivism (within-subject factor) are different across the cultures (between subject factor) while taking into account response tendencies of participants. Finally, rank orders might be compared across cultures if researchers are interested in the relative order of constructs (e.g., Church & Katigbak, 1992).

In contrast to ipsative scores produced by within-subject and double standardization, within-group or within-culture standardization can be used for correlational techniques such as regression (or factor analysis). However, the meaning of resulting coefficients might change and researchers have to be aware of the correct interpretation of regression weights as discussed above. For mean comparisons, again, the meaning of the resulting scores changes depending on the standardization procedure and the number of involved variables used.

Finally, partialing the mean across items as recommended by Schwartz (1992) might be appropriate when used with the Schwartz Value Survey. Researchers using this technique for other instruments should consider limitations stemming from the potential problem of singularity, multicollinearity, and differential adjustment of scales.

The previous points do not imply that researchers should refrain from using Likert-type response scales, nor should researchers avoid standardization. The main point of this article is to highlight some of the implications of standardization in terms of statistical properties as well as the theoretical interpretation of results. Adjustments are sometimes needed and can provide useful information that would be obscured if no standardization was used (e.g., Raudenbush & Bryk, 2002). The crucial question is for what purposes researchers want to standardize raw data. In my opinion, it is necessary to distinguish between standardization used for (a) examining structural relationships among variables or theoretical constructs (e.g., dimensionality of constructs) and (b) explaining level differences between groups (Van de Vijver & Leung, 1997). Concerning the examination of structural relationships between variables, I would suggest caution when using ipsatized data in factor analytical designs. Multidimensional scaling might be more appropriate for ipsative data, as well as in cases where concerns exist that response biases might obscure structural relationships. Explaining level differences between groups—for example, through the use of *t* tests, ANOVAs, regression, or hierarchical linear modeling—sometimes requires standardization techniques and it is important that researchers are aware of the appropriate interpretation.

In this article, I was only concerned with the methodological and statistical implications of standardization when trying to eliminate response bias. I did not address the issue of whether response bias is a factor that needs to be controlled for. Increasingly, evidence becomes available that response bias might actually be a variable of substantive interest and a true indicator of cross-cultural differences. For example, Smith (2004) and Van Hemert et al. (2002) pointed out that cross-cultural differences in response bias can be explained in terms of psychological dimensions, such as the ones derived from Hofstede's (1980) work.

Therefore, these differences might better reflect different communication styles across cultures rather than bias that needs to be controlled for. The focus of the investigation might be important for how individual researchers deal with these issues. More research is needed to explore this possibility. In the meantime, appropriate statistics are available that can detect item bias (Cheung & Rensvold, 2000) or differential item functioning (Van de Vijver & Leung, 1997) across different cultural samples.

NOTE

1. The nature of these group differences is not of interest for this example. These differences may represent true group level differences, some measurement deficiencies of self-esteem at a group level, some sort of response bias, omitted variables at a group level that correlate highly with self-esteem, and so forth.

REFERENCES

Adams, F. M., & Osgood, C. E. (1973). A cross-cultural study of the affective meaning of color. *Journal of Cross-Cultural Psychology*, *4*, 135-155.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Aycan, Z., Kanungo, R. N., & Sinha, J. P. (1999). Organizational culture and human resource management practices: The model of culture fit. *Journal of Cross-Cultural Psychology*, *30*, 501-526.

Baron, H. (1996). Strengths and limitations of ipsative measurement. *Journal of Occupational and Organizational Psychology*, *69*, 49-56.

Bartram, D. (1996). The relationship between ipsatized and normative measures of personality. *Journal of Occupational and Organizational Psychology*, *69*, 25-39.

Boehnke, K., Silbereisen, R. K., Eisenberg, N., Reykoowski, J., & Palmonari, A. (1989). Developmental patterns of prosocial motivation: A cross-national study. *Journal of Cross-Cultural Psychology*, *20*, 219-243.

Bond, M. H. (1988). Finding universal dimensions of individual variation in multicultural studies of values: The Rokeach and Chinese values surveys. *Journal of Personality and Social Psychology*, *55*, 1009-1015.

Bond, M. H., Wan, K., Leung, K., & Giacalone, R. A. (1985). How are responses to verbal insult related to cultural collectivism and power distance? *Journal of Cross-Cultural Psychology*, *16*, 111-126.

Butcher, J. N., Lim, J., & Nezami, E. (1998). Objective study of abnormal personality in cross-cultural settings: The Minnesota Multiphasic Personality Inventory (MMPI-2). *Journal of Cross-Cultural Psychology*, *29*, 189-211.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumptions of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-Cultural Psychology*, *30*, 555-574.

Chan, W. (2003). Analyzing ipsative data in psychological research. *Behaviormetrika*, *30*, 99-121.

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*, 160-186.

Chinese Culture Connection. (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology*, *18*, 143-164.

Chun, K.-T., Campbell, J. B., & Yoo, J. H. (1974). Extreme response style in cross-cultural research: A reminder. *Journal of Cross-Cultural Psychology*, *5*, 465-479.

Church, A. T., & Katigbak, M. S. (1992). The cultural context of academic motives: A comparison of Filipino and American college students. *Journal of Cross-Cultural Psychology*, *23*, 40-58.

Closs, S. J. (1996). On the factoring and interpretation of ipsative data. *Journal of Occupational and Organizational Psychology*, *69*, 41-47.

Cornwell, J. M., & Dunlop, W. P. (1994). On the questionable soundness of factoring ipsative data: A response to Saville and Willson (1991). *Journal of Occupational and Organizational Psychology*, *67*, 89-100.

Crittenden, K. S., Fugita, S. S., Bae, H., Lamug, C. B., & Lin, C. (1992). A cross-cultural study of self-report depressive symptoms among college students. *Journal of Cross-Cultural Psychology*, *23*, 163-178.

Everitt, B. S., & Wykes, T. (1999). *A dictionary of statistics for psychologists*. London: Arnold Publishers.

Feather, N. T., Volkmer, R. E., & McKee, I. R. (1992). A comparative study of the value priorities of Australians, Australian Baha'is and expatriate Iranian Baha'is. *Journal of Cross-Cultural Psychology*, *23*, 95-106.

Field, A. (2000). *Discovering statistics using SPSS for Windows*. London: Sage.

Fioravanti, M., Gough, H. G., & Frere, L. J. (1981). English, French, and Italian adjective check lists: A social desirability analysis. *Journal of Cross-Cultural Psychology*, *12*, 461-472.

Fogarty, G. J., & White, C. (1994). Differences between values of Australian Aboriginal and non-Aboriginal students. *Journal of Cross-Cultural Psychology*, *25*, 394-408.

Goodwin, R., Nisharadze, G., Kosa, L. A. N. L. E., & Emelyanova, T. (1999). Glasnost and the art of conversation: A multilevel analysis of intimate disclosure across three former Communist cultures. *Journal of Cross-Cultural Psychology*, *30*, 72-90.

Grimm, S., Church, A. T., Katigbak, M. S., & Reyes, J. A. S. (1999). Self-described traits, values and moods associated with individualism and collectivism: Testing I-C theory in an individualistic (US) and collectivistic (Philippine) culture. *Journal of Cross-Cultural Psychology*, *30*, 466-500.

Gudykunst, W. B., Gao, G., Schmidt, K. L., Nishida, T., Bond, M. H., Leung, K., et al. (1992). The influence of individualism-collectivism, self-monitoring, and predicted-outcome value on communication in ingroup and outgroup relationships. *Journal of Cross-Cultural Psychology*, *23*, 196-213.

Harrington, L., & Liu, J. H. (2002). Self-enhancement and attitudes toward high achievers: A bicultural view of the independent and interdependent self. *Journal of Cross-Cultural Psychology*, *33*, 37-55.

Hasegawa, T., & Gudykunst, W. B. (1998). Silence in Japan and the United States. *Journal of Cross-Cultural Psychology*, *29*, 668-684.

Hicks, L. E. (1970). Some properties of ipsative, normative and forced-choice normative measures. *Psychological Bulletin*, *74*, 167-184.

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Hofstee, W. K. B., Ten Berge, J. M. F., & Hendriks, A. A. J. (1998). How to score questionnaires. *Personality and Individual Differences*, *25*, 897-909.

Howell, D. C. (1997). *Statistical methods for psychology*. Belmont, CA: Duxbury Press.

Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO personality inventory. *Journal of Cross-Cultural Psychology*, *28*, 192-218.

Hui, C. H., & Yee, C. (1994). The shortened Individualism-Collectivism scale: Its relationship to Demographic and work-related variables, *Journal of Research in Personality*, *28*, 409-424.

Johnson, C. E., Wood, R., & Blinkhorn, S. F. (1988). Spuriouser and spuriouser: The use of ipsative personality tests. *Journal of Occupational Psychology*, *63*, 153-162.

Kashima, Y., Siegal, M., Tanaka, K., & Kashima, E. S. (1992). Do people believe behaviors are consistent with attitudes? Towards a cultural psychology of attribution processes. *British Journal of Social Psychology*, *31*, 111-124.

Kirkman, B. L., & Shapiro, D. L. (2001). The impact of team members' cultural values on productivity, cooperation and empowerment in self-managing work teams. *Journal of Cross-Cultural Psychology*, *32*, 597-617.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling* (Sage University Paper No. 07). London: Sage.

Kurman, J., & Sriram, N. (1997). Self-enhancement, generality of self-evaluations, and affectivity in Israel and Singapore. *Journal of Cross-Cultural Psychology*, *28*, 421-441.

Landis, D., & O'Shea, W. A. (2000). Cross-cultural aspects of passionate love: An individual differences analysis. *Journal of Cross-Cultural Psychology*, *31*, 752-777.

Leung, K., & Bond, M. H. (1989). On the empirical identification of dimensions for cross-cultural comparisons. *Journal of Cross-Cultural Psychology*, *20*, 133-151.

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

Little, T. D. (2000). On the comparability of constructs in cross-cultural research: A critique of Cheung and Rensvold. *Journal of Cross-Cultural Psychology*, *31*, 213-219.

Marshall, R. (1997). Variances in levels of individualism across two cultures and three social classes. *Journal of Cross-Cultural Psychology*, *28*, 490-495.

Matsumoto, D., Weissman, M. D., Preston, K., Brown, B. R., & Kupperbusch, C. (1997). Context-specific measurement of individualism-collectivism on the individual level: The individualism-collectivism interpersonal assessment inventory. *Journal of Cross-Cultural Psychology*, *28*, 743-767.

Merrit, A. (2000). Culture in the cockpit: Do Hofstede's dimensions replicate? *Journal of Cross-Cultural Psychology*, *31*, 283-301.

Minsel, B., Becker, P., & Korchin, S. J. (1991). A cross-cultural view of positive mental health: Two orthogonal main factors replicable in four countries. *Journal of Cross-Cultural Psychology*, *22*, 157-181.

Munroe, D. (1979). Locus of control attribution: Factors among Blacks and Whites in Africa. *Journal of Cross-Cultural Psychology*, *10*, 157-172.

Nelson, M. R., & Shavitt, S. (2002). Horizontal and vertical individualism and achievement values: A multimethod examination of Denmark and the United States. *Journal of Cross-Cultural Psychology*, *33*, 439-458.

Ohbuchi, K. I., Fukushima, O., & Tedeschi, J. T. (1999). Cultural values in conflict management: Goal orientation, goal attainment, and tactical decisions. *Journal of Cross-Cultural Psychology*, *30*, 51-71.

Phalet, K., & Claeys, W. (1993). A comparative study of Turkish and Belgian youth. *Journal of Cross-Cultural Psychology*, *24*, 319-343.

Ralston, D. A., Vollmer, G. R., Srinvasan, N., Nicholson, J. D., Tang, M., & Wan, P. (2001). Strategies of upward influence: A study of six cultures from Europe, Asian and America. *Journal of Cross-Cultural Psychology*, *32*, 728-735.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.
Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology*, *37*, 345-356.
Schaffer, D. R., Crepaz, N., & Sun, C. R. (2000). Physical attractiveness stereotyping in cross-cultural perspective: Similarities and differences between Americans and Taiwanese. *Journal of Cross-Cultural Psychology*, *31*, 557-582.
Schultz, P. W., & Zelezny, L. C. (1998). Values and proenvironmental behavior: A five-country survey. *Journal of Cross-Cultural Psychology*, *29*, 540-558.
Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1-65.
Schwartz, S. H. (1994). Beyond individualism-collectivism: New cultural dimensions of values. In U. Kim, H. C. Triandis, C. Kagitcibasi, S. C. Choi, & G. Yoon (Eds.), *Individualism and collectivism: Theory, method and applications* (pp. 85-119). Thousand Oaks, CA: Sage.
Seki, K., Matsumoto, D., & Imahori, T. T. (2002). The conceptualisation and expression of intimacy in Japan and the United States. *Journal of Cross-Cultural Psychology*, *33*, 303-319.
Singelis, T. M., Bond, M. H., Sharkey, W. F., & Lai, C. S. Y. (1999). Unpackaging cultures's influence on self-esteem and embarrassability. *Journal of Cross-Cultural Psychology*, *30*, 315-341.
Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, *35*, 50-61.
Smith, P. B., Peterson, M. F., & Schwartz, S. H. (2002). Cultural values, sources of guidance and their relevance to managerial behaviors: A 47 nation study. *Journal of Cross-Cultural Psychology*, *33*, 188-208.
Stewart, S. M., Bond, M. H., Deeds, O., & Chung, S. F. (1999). Intergenerational patterns of values and autonomy expectations in cultures of relatedness and separateness. *Journal of Cross-Cultural Psychology*, *30*, 575-593.
Tenopyr, M. L. (1988). Artefactual reliability of forced-choice scales. *Journal of Applied Psychology*, *73*, 749-751.
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., McCusker, C., Betancourt, H., Iwao, S., Leung, K., Salazar, J. M., et al. (1993). An etic-emic analysis of individualism and collectivism. *Journal of Cross-Cultural Psychology*, *24*, 366-383.
Van de Vijver, F., & Leung, K. (1997). *Methods and data analysis of comparative research*. Thousand Oaks, CA: Sage.
Van de Vijver, F. J. R., & Poortinga, Y. H. (2002) Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology*, *33*, 141-156.
Van de Vliert, E., & Janssen, O. (2002). “Better” performance motives as roots of satisfaction across more and less developed countries. *Journal of Cross-Cultural Psychology*, *33*, 380-397.
Van Hemert, D. A., Van de Vijver, F. J. R., Poortinga, Y. H., & Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences*, *33*, 1229-1249.
Wagner, W., Kirchler, E., Clack, F., Tekarslan, E., & Verma, J. (1990). Male dominance, role segregation, and spouses' interdependence in conflict: A cross-cultural study. *Journal of Cross-Cultural Psychology*, *21*, 48-70.
Watkins, D., & Cheung, S. (1995). Culture, gender and response bias: An analysis of responses to the self-description questionnaire. *Journal of Cross-Cultural Psychology*, *26*, 490-504.
Yamaguchi, S., Kuhlman, D. M., & Sugimori, S. (1995). Personality correlates of allocentric tendencies in individualist and collectivist cultures. *Journal of Cross-Cultural Psychology*, *26*, 658-672.

Ronald Fischer is a lecturer in psychology at Victoria University, Wellington, New Zealand. He received his D.Phil from the University of Sussex, UK. Through traveling, he got interested in culture and he began to study cultural aspects of social and organizational psychology, especially justice perceptions, culture and attitudes, and behavior, as well as research methods and statistics for cross-cultural investigations.