whereas predictors based on interval data are usually analyzed with a regression model. Yet, the choice is more a matter of convenience than of principle, given the close link between the two. Cohen and Cohen (1983) and Pedhazur (1982) describe how an analysis of variance can be seen as a regression analysis; the independent variables of an analysis of variance are the predictors of a regression analysis. The significance tests of the regression coefficients in the regression analysis (which are t tests) yield similar results as the significance tests of the analysis of variance (F test). Specifically, the squared t values of the regression statistics are equal to the F ratios of the analysis of variance.

Regression analyses can be used to identify relationships that hold both within and across cultures, which are actually structure-oriented issues. This approach is illustrated with two variables (X and Y). The first step in this technique is to obtain a pan-cultural regression equation of Y on X, in which data from all cultures are included. In the second step, culture is added as a dummy variable, and another regression analysis including X, the dummy variable, and the interaction of X and the dummy variables as predictors, is carried out. The multiple correlations of the two equations are then tested for equality. If there is no significant difference between these two multiple correlations, it is concluded that the relationship holds within each culture and across all cultures. In other words, the regression weights and the intercepts of the equation are similar in all cultures. A significant difference of the two multiple correlations points to the presence of intergroup differences on the dependent variable not explained by X. For an elaboration of this approach, see Leung (1987) and Poortinga and Van de Vijver (1987).

When more than one dependent variable is involved, covariance structure analysis employing models such as LISREL and EQS becomes appropriate. The basic idea is to test whether the interrelationships of the variables are similar in different cultures. For instance, confirmatory factor analysis involving two or more cultures is frequently conducted to see if a factor structure is similar in different cultures (e.g., De Groot, Koot, & Verhulst, 1994; Leung et al., 1992; Marsh & Byrne, 1993; Sachs, 1992; Watkins, 1989). The reader is referred to Poon, Chan, Lee, and Leung (1993) for a method to compare the equivalence of a covariance structure in a large number of cultures.

Multilevel analysis is another procedure that can be used to address both level- and structure-oriented issues (Bock, 1989; Bryk & Raudenbush, 1992; Goldstein, 1987; Lee, 1990). Even though these procedures have not been applied in cross-cultural research, their potential value is obvious. At least two levels of analysis are possible in cross-cultural research (e.g., Leung, 1989; Leung & Bond, 1989). In the culture-level approach, culture is the unit of analysis, and the results obtained are characterizations of cultures, but not individuals. The classic study on values by Hofstede (1980) is based on this approach. There is no assumption with regard to whether relationships found across cultures will hold within each of the cultures included in the analysis. Culture-level analyses can guard against the ecological fallacy, the incorrect application of culture-level characteristics to individuals. When a culture is known to be individualistic, the mean score on a scale of individualism will be higher than in a collectivistic culture, but it does not mean

that each person has a high score on individualism. Furthermore, cross-level inferences can be fallacious because of a difference in meaning of constructs at individual and cultural levels. Gender at the individual level can have two values, male and female; an aggregated gender score at the cultural level refers to the proportion of males and females in a group, which is quite a different concept.

In the individual-level approach, the individual is the unit of analysis, and this is the dominant approach in cross-cultural psychology. The relationships between variables at individual and cultural levels need not be equal (cf. Ostroff, 1993). Yet, it is more elegant theoretically to demonstrate their equality. An example can be found in "subsystem validation," in which "hypotheses are examined both intraculturally and cross-culturally, so that explanatory variables may be tested at two levels" (Berry & Dasen, 1974, p. 19). The objective of this approach is to establish that the relationships among a set of variables hold within a culture as well as across cultures. For instance, in the classic study by Segall, Campbell, and Herskovits (1966), it was found that when a culture is associated with a more "carpentered" environment (more corners formed of intersecting planes perpendicular to each other), people from this culture are more susceptible to geometric illusions. This finding explains why one cultural group is more susceptible to geometric illusions than another cultural group. Their findings also imply that if a person is exposed to a more carpentered environment, he or she is more susceptible to geometric illusions. Thus, their findings are able to explain cultural differences as well as individual differences in susceptibility to geometric illusions.

Statistical Tests of Cross-Cultural Differences: Structure-Oriented Techniques

Cross-cultural psychologists are often interested in a comparison of the *structure* underlying the data rather than a direct comparison of the observed variables as discussed in the previous section. For instance, much research has been devoted to the question of whether the structure of intelligence is universal. Do the same cognitive processes contribute to test performance in different cultural groups? These questions have probably their intellectual roots in the notion of the psychic unity of humankind (Tylor, 1871), which can be interpreted as the idea that the structure behind human behavior is universal.

While multivariate statistical techniques are often applied in structure-oriented analysis, ANOVA or t tests are commonly applied when the independent variables are discrete. The focus is here to evaluate whether the differences of the dependent variable across the various levels of the independent variable are similar or different in each culture. For instance, Buss (1989) applied t tests to examine mate preferences of males and females in 33 cultures. He confirmed hypothesized gender differences in mate preferences in the cultures studied.

The similarity of psychological structure has been studied mostly by means of factor analysis (Harman, 1976; McDonald, 1985). Like regression analysis, factor analysis postulates that an observed score is a weighted sum of a usually limited set of contributing factors. Unlike in regression analysis, however, the

contributing factors are not observable in factor analysis. An observed score, for example, an intelligence test score of a person, is a weighted sum of unobservable factor scores, such as reasoning ability, perceptual speed, and memory, which in turn are determined by subtests of the intelligent test. Based on the intercorrelations of the subtests, factor analysis determines the score of each person on the factors and the correlations of the subtests with the factors, the so-called factor loadings. We will not dwell upon the classical problems of factor analysis here, which include the determination of the number of factors and the rotation problem, because these problems are inherent to factor analysis and not unique to its cross-cultural applications.

When factor analysis is applied to cross-cultural data, one major question to be considered is the (lack of) similarity of the factor analytic solution across the groups. The question amounts to a check on the equality of the factor loadings. Do the instruments (tests, items, and observational measures) have the same correlations with the factors in each cultural group? The equality of the factor loadings is sometimes visually checked, which is a questionable practice as more powerful procedures exist.

Such a procedure starts with a so-called target rotation (e.g., McDonald, 1985). Factor analytic solutions can be freely rotated (the rotation problem). This subjectivity is usually "solved" by applying a rotation procedure such as Varimax. However, independently obtained factor loadings (no matter whether they are rotated by Varimax or any other rotation procedure) may be more similar than a visual inspection may suggest. The factor loading matrices may be rotated to each other in order to maximize their agreement. This is a legitimate procedure because of the arbitrariness of factor analytic solutions. In a target rotation the axes are rotated in such a way that the agreement between the sets of factor loadings is optimized. One of the groups is arbitrarily chosen as the target to which the factor loading matrices of the other groups will be rotated.

After having rotated the factor loadings, their similarity can be evaluated in a factor by factor comparison by means of a coefficient of agreement. The most often used coefficient of agreement has been developed by Tucker (1951); it has become known as *Tucker's coefficient of agreement* and also as proportionality coefficient (Zegers & Ten Berge, 1985). The coefficient is comparable to a correlation coefficient, the only difference is that, unlike a correlation coefficient, the coefficient of agreement is sensitive to a constant that is added to one of the variables. As an alternative to the coefficient of agreement, the identity coefficient can be proposed that is sensitive to any linear transformation (Zegers & Ten Berge, 1985). The coefficient is defined as

$$e_{xy} = \frac{2 \sum x_i y_i}{\sum x_i^2 + \sum y_i^2},$$

in which x_i and y_i represent the factor loadings in the two groups. As a rule of thumb, values of the identity coefficient lower than .90 are taken to point to a lack of agreement and values higher than .95 are seen as evidence for the similarity of the factor matrices.

Other procedures have also been developed to evaluate the agreement of factor analytic solutions across groups. Thus, equivalence of the Eysenck personality scales (e.g., Eaves, Eysenck, & Martin, 1989; Eysenck & Eysenck, 1983) has often been studied employing a procedure proposed by Kaiser, Hunka, and Bianchini (1971). There has been some debate as to whether procedures such as those proposed by Kaiser et al. (1971) or by Tucker (1951) are sufficiently powerful to detect item bias. Using simulated data, the critics (e.g., Bijnen, Van der Net, & Poortinga, 1986; Van de Vijver & Poortinga, 1994) have shown that values well over .90 can be obtained when in fact there are items with dissimilar loadings across groups. So, caution is required because these agreement indices sometimes do not reflect the influence of nonequivalent items.

Multidimensional scaling procedures have also been employed to compare the structure of cross-cultural data sets. An example can be found in the work of Schwartz (e.g., Schwartz, 1992, 1994; see also Endler, Lobel, Parker, & Schmitz, 1991; Russell, Lewicka, & Niit, 1989). Multidimensional scaling attempts to reproduce the distances between stimuli (such as test or item scores or behavioral measures) in a small number of dimensions. To compare multidimensional scaling solutions obtained in different groups, the technique has the same rotational problem as factor analysis. Distances between stimuli are not affected by (orthogonal) rotations of the axes. Consequently, configurations of such analyses as obtained in different cultural groups have an arbitrary spatial orientation. Target rotations have to be applied prior to an evaluation of the agreement of the solutions; when such rotations are not carried out, the agreement will be underestimated. The procedure of carrying out target rotations and computing an index of agreement of the solutions described above for factor analysis also applies here. No empirical applications of target rotations following a multidimensional scaling procedure are known to the authors.

Cluster analysis is another technique that aims to reduce a large set of correlations or distances to a smaller number of dimensions or factors. However, although this technique is suitable for cross-cultural research, it has not been used much in the cross-cultural literature.

The final techniques to be discussed here have also been mentioned in the discussion of level-oriented techniques, namely the analysis of covariance structures or linear structure models, which analyzes the covariance matrix of a set of measures (Bollen, 1989; Byrne, 1989). Two common computer packages for covariance modelling, namely LISREL (Byrne, 1989; Jöreskog & Sörbom, 1993) and EQS (Bentler, 1992; Byrne, 1994), can be used to analyze multigroup data. Several applications of linear structure models in cross-cultural research can be envisaged, including confirmatory factor analysis and the analysis of multitrait–multimethod (or monotrait–multitrait) matrices for assessing method bias. Confirmatory factor analysis is a versatile tool for testing cross-cultural differences in covariance structures. Compared to the classical factor analytic procedures described above, confirmatory factor analysis allows for the test of a large set of hierarchically linked hypotheses of cross-cultural invariance. The analyses usually consist of two series. The first analysis tests whether the covariance matrix of the measures is the same for all cultural groups. Fit tests play an essential role in

covariance structure analysis. Unfortunately, fit tests in LISREL and EQS are not easy to interpret. The overall goodness-of-fit index is a chi-square distributed variable that is known to be sensitive to sample size; in large samples small intergroup differences in the covariance matrix will yield a significant value. Various fit measures have been developed that are less dependent on sample size (Bollen & Long, 1993).

If the null hypothesis is not rejected, it is highly likely that the psychological structure underlying the performance is identical across cultural groups. If the hypothesis of equal covariance matrices has to be rejected (which is usually the case), the second series of hypothesis testing will start. The second series consists of a set of hierarchically nested models that successively increase the number of equality constraints across groups. The choice of the models is free, but the order specified here (as well as in various other sources; e.g., Jöreskog & Sörbom, 1993; Vandenberg & Self, 1993; Van de Vijver & Harsveld, 1994) usually follows a theoretically relevant sequence. The first analysis of the second series specifies an equal number of factors in each group. The specification of the number of factors should be based on preliminary analyses or on earlier research findings. If the hypothesis of an equal number of factors has to be rejected, an exploratory factor analysis can be carried out to investigate the reason for the lack of fit; for instance, factors may have split up or merged in various cultural groups. If this is the case, the analyses may proceed with different numbers of factors across groups. A total lack of correspondence of the factors, after the possibilities of split and merged factors have been explored, would point to a small overlap in the psychological meaning of the instrument across the groups. Such a lack of correspondence can be expected for statistical reasons when many correlations in one or more groups are close to zero. The input variables of the factor analysis (item scores, test scores, etc.) are then largely independent of each other and the use of any multivariate technique should be strongly questioned.

The second step of the second series will test whether the matrix of factor loadings can be considered equal across the cultural groups. A set of factor loadings have to be left free in the first group; the values in the other groups are constrained to have the same values as the factor loadings in the first group. This will again yield various fit indices, among which are incremental fit indices. Because this model is subsumed in the previous model (stating an equal number of factors), the difference in the chi-square fit indices of the two models can be interpreted meaningfully: the difference in chi-square values follows a chi-square distribution with a number of degrees of freedom that is equal to the difference of the number of degrees of freedom of the two models. Acceptance of the hypothesis points to the equality of the composition of the latent factors across the groups. Rejection of the hypothesis provides evidence that the psychological structure underlying the data is dissimilar across the cultural groups. A better fit may be found when only a subset of the factor loadings are set equal to each other across cultural groups. If this is the case, subtle differences in the psychological structure have been observed. If an acceptable fit of a model specifying equal factor loadings is found, constraints can be added, such as equality of the covariance

matrices of the latent factors in all groups. The study of the fit of hierarchically nested models provides a flexible tool to analyze covariance structures. It can be used to detect smaller or larger differences in psychological meaning of measurement instruments across cultural groups. For an example of confirmatory factor analysis in cross-cultural psychology, the reader is referred to Watkins (1989).

Covariance structure analysis can also be employed for causal modelling. A set of variables, either consisting of observed variables or of a combination of observed and latent (i.e., unobservable) variables, are assumed to have *a priori* specified antecedent–consequence relationships. The model can be based on theoretical expectations or an earlier exploratory study. In some cases the exploratory study and the test of the causal model are derived from random splits of the same sample; the model is then developed on half of the data and cross-validated in the other half. An example of a cross-cultural application of a causal model can be found in Van Haaften and Van de Vijver (in press). The number of applications of causal modelling in the cross-cultural literature is limited (cf., however, Little, Oettingen, & Baltes, 1995; Little, Oettingen, Stetsenko, & Baltes, 1995), but given its flexibility and usefulness, its use is recommended.

Covariance structure analysis can also be used to assess method bias, either in a test–retest design or in a monotrait–multimethod approach. In the latter case all methods that are employed (observed variables) load on the same latent factor(s). LISREL and EQS allow for a test of the similarity of the loadings of the methods across the two groups.

Finally, it should be pointed out that the distinction between level- and structure-oriented techniques is not strict in some statistical techniques. In regression, multilevel, and covariance modelling techniques, the differentiation between level- and structure-oriented questions is quite subtle. For instance, suppose that educational achievement is predicted on the basis of a set of aptitude tests in two different cultures and equality of regression lines is tested. Similarity of regression coefficients involves structural relationships whereas equality of the intercept would refer to level-oriented relationships. The same applies to multi-level models that can tackle both level- and structure-oriented questions. In empirical applications of covariance modelling there tends to be an emphasis on structural relationships. However, the models are sufficiently flexible to deal with intergroup differences in averages as well. In sum, the designation of regression and multilevel models as level-oriented and covariance modelling as structure-oriented is more inspired by their common usage than by theoretical characteristics of these models. They could as well be seen as hybrid models.

Four Common Types of Comparative Studies

In the remainder of the chapter we shall make a distinction between four types of cross-cultural studies, depending on whether the orientation is exploratory or hypothesis testing, and on whether or not contextual factors are considered (see Table 7–2).

TABLE 7–2 Common types of comparative studies

	Orientation	
Consideration of contextual factors	Hypothesis testing	Exploration
No	Generalizability	Psychological differences
Yes	Theory-driven	External validation

The first two types emphasize hypothesis testing. The first kind of studies, *generalizability studies*, attempts to establish the generalizability of research findings obtained in one, typically Western, group to other Western or non-Western groups. In general, these studies make little or no reference to local cultural elements.

In the second type, called *theory-driven studies*, cultural factors are part of the theoretical framework. Cultural variation is deliberately sought as a validation of the model, and specific *a priori* predictions are proposed and tested. The framework is tested by sampling various cultures that differ on some focal dimension. Theory-driven studies test a theory about a particular relationship between cultural variables and a psychological outcome. Contextual elements are crucial in this type of studies.

Hypothesis testing receives little emphasis in the following two types of cross-cultural research. The first type, *psychological differences* studies, is probably most common in the literature. A measurement instrument is applied in at least two cultures and the researcher is interested in whether there are any differences in averages, standard deviations, reliabilities, or other psychometric properties of the instrument across the cultural groups. Usually, the original instrument has been applied before in a Western context, and an application of the instrument in another cultural group is thought to provide an interesting extension. There is often no compelling theory about the nature of the cross-cultural differences to be expected. Contextual factors are typically not included in the design, and post hoc explanations are invoked to interpret the cross-cultural differences observed.

The last type of cross-cultural research refers to what has been called *external validation*, which attempts to explore the meaning and causes of cross-cultural differences with the aid of contextual factors. In this type of studies, specific *a priori* hypotheses are absent and usually a large set of contextual variables are included in an exploratory manner. Only a few statistical techniques have been applied in external validation studies. Regression analysis is the most frequently applied technique, which assesses the effectiveness of independent variables in explaining cross-cultural variations in the dependent variable. This kind of validation does not address structural or scalar equivalence, but aims at providing evidence for a particular interpretation of cross-cultural differences.

Poortinga and Van de Vijver (1987) have outlined a general procedure for external validation with the inclusion of covariates. The procedure presupposes

that data (tests, observational instruments, interviews, surveys, etc.) are collected in at least two cultural groups. Data should also be collected on additional variables, termed *context variables*, that are likely to be able to explain cross-cultural differences that may be obtained. The data analysis starts with an analysis of variance to test the null hypothesis of no cultural differences. In the next analysis context variables are introduced; they are used as covariates in an analysis of covariance or as predictors in a regression analysis. In terms of an analysis of variance, the main effect of culture is tested twice; the first analysis tests group differences before correction for context variables; the second analysis tests intergroup differences in residual scores after orrection. Let us call the corresponding F ratios F_1 and F_2, respectively. Significant F values point to intergroup differences. A comparison of the significance of F_1 and F_2 can yield various possibilities. If F_1 is not significant, there are no intergroup differences to be explained (even though there is still a remote possibility that the introduction of context variables could reveal significant intergroup differences). Context variables will play a central role when F_1 is significant. Introduction of context variables can give rise to three possibilities. First, context variables may be unrelated to the dependent variable, in which case intergroup differences cannot be accounted for by these context variables. Second, context variables can be related to the dependent variable and intergroup differences on the dependent variable become smaller after correction, but they are still significant. In this case context variables provide a partial explanation of intergroup differences. Third, when F_2 is not significant, intergroup differences can be accounted for entirely by the context variables.

It should be pointed out that internal and external validation procedures have different goals. Internal validation aims at establishing the cross-cultural equivalence of the data. The key question is to ascertain whether the scores of individuals in all cultural groups can be directly compared. In external validation procedures, scalar equivalence is assumed, and the research goal is to shed light on the meaning and interpretation of the cross-cultural differences. In other words, internal validation procedures attempt to detect and remove culturally biased items, whereas external validation procedures attempt to explore the causes of cross-cultural differences observed.

Methods and Analysis of Four Common Types of Comparative Studies

In this section, the four types of cross-cultural studies—generalizability, theory-driven, psychological differences, and external validation—are examined with regard to the following issues: sampling of cultures and of subjects, procedure, design, analysis, and major strengths and weaknesses. See Table 7–3 for a summary. Examples from the literature will be described.

In *generalizability studies* a theory, a correlational or causal relationship, or an instrument derived from a theory is tested in another cultural context. The goal of the study is to establish the generalizability of the theory, the relationship, or

TABLE 7–3 Methods and analysis for the four common types of comparative studies

Type of study	Sampling of cultures	Design	Major analysis	Major strength	Major weakness
Generalizability	convenience	replication of original study or new study	structure techniques (e.g., correlations, factor analysis, analysis of co-variance structures)	study of equivalence	no contextual variables included
Psychological differences	systematic or convenience sampling	replication of original study or new study	both level (e.g., *t* test and ANOVA) and structure techniques	"openmindedness" about cross-cultural differences	ambiguous interpretation
Theory-driven	systematic (maximize contrast on focal variable)	new study; covariates may be included	both level and structure techniques	study of relationship of cultural factors and behavior	lack of attention to alternative interpretations
External validation	systematic	measures at different levels of aggregation; covariates included	level techniques	focus on interpretation of cross-cultural differences	choice of covariate variables may be meaningless

the instrument. The cultures are often chosen on the basis of convenience sampling. Two different subject sampling schemes can be applied: random or matched sampling. Generalizability will be high when the original results are replicated and the subjects are sampled randomly. However, a lack of replicability cannot be interpreted unambiguously in random sampling. Negative findings could be due to cultural differences or to a lack of equivalence of the samples. A new data set using matched samples is then required to establish a more unambiguous interpretation. The procedure of the study usually follows the procedure of the original one; in some cases stimuli may be added to enhance the appropriateness of the instrument for the local context. The design, too, is a replication of the original one.

For replications, data analysis will consist of two parts: the first part will be identical to the analysis of the original study. Second, because the establishment of generalizability is the aim of the study, an assessment of the similarity of the original and new results is required. Factor analyses, followed by target rotation and the computation of an index of agreement or multigroup analyses of covariance structures, are to be preferred over a more informal evaluation of the similarity of the outcome. Compared to studies in which results can be constrasted with those obtained in previous studies conducted in other cultures, studies that are conducted simultaneously in a number of cultures will employ more exploratory data analyses for identifying cultural similarities and differences in the results.

The major strength of generalizability studies is their ability to test the equivalence of the results across cultures. As prior data are available with which new data sets can be compared, various hypotheses about cross-cultural differences and similarities can be investigated. A weakness of generalizability studies is that they often fail to include contextual variables. If cultural differences are found, it is often not at all clear how these should be interpreted. Furthermore, bias analyses are infrequently carried out in these studies. Thus, it is too common to take unexpected differences in item scores at face value (instead of carrying out an item bias analysis).

Most examples of generalizability studies in the literature involve studies of applications of an instrument, derived from a theory. Schwartz (1992) has collected data from various countries on the universality of the structure of human values. Irvine (1979) and Vernon (1969, 1979) have compared the structure of intelligence across cultures. A study of the choice of conflict resolution procedures (Leung, 1987) is an example of a cross-cultural study of a causal relationship. Amir and Sharon's (1987) replication of Western social–psychological studies in Israel among high school and university students is another good example of a generalizability study. Finally, there are many attempts recently to validate the big-five personality factors in a variety of cultures (e.g., McCrae & Costa, 1985; McCrae & John, 1992).

In *theory-driven studies*, cultures are often systematically sampled in order to maximize their contrast on some focal variable. The sampling of subjects requires scrutiny. The cultures in the sample will often differ in many more respects than

those intended and of interest to the researcher. As the matching of the groups on all relevant ambient variables cannot be achieved, contextual measures should be added to enhance the interpretability of the findings. The measurement instruments should assess various other variables on which the cultures differ and which could obscure the cross-cultural differences being studied. The experimental procedure used is often similar across cultures. Because theory-driven studies are usually level-oriented studies, data analysis usually involves analysis of variance or covariance. In the latter case the context variables are the covariates.

The most important strength of theory-driven studies is the explicit postulation of the relationship between cultural factors and the focal behavior, which is often considered the main goal of cross-cultural psychology (e.g., Berry, Poortinga, Segall, & Dasen, 1992). The major weakness of theory-driven studies is their lack of attention to item bias and alternative explanations for the cross-cultural differences observed.

An example of a theory-driven study without covariates is Berry's (1976; Berry et al., 1986) study of the cognitive styles of hunters and food gatherers. Cultural variations in perceptual styles, educational patterns, and societal structures are all hypothesized to be interrelated and to be functionally related to the food gathering patterns of a cultural group. An example with a covariate is Earley's (1989) study in which American subjects were found to show more social loafing (the phenomenon that people work less when they are in a group than when they have to do the same task individually) than Chinese subjects. In the study, subjects' individualism–collectivism score was measured as a covariate. After controlling for cross-cultural differences in individualism–collectivism in an analysis of covariance, the cross-cultural differences in social loafing disappeared. The covariance analysis provided strong evidence for the role of individualism–collectivism in explaining cross-cultural differences in social loafing.

Studies of *psychological differences* involve the application of a measurement instrument such as a test, an interview scheme, or an observation scale, in a new cultural context. The purpose is to explore cross-cultural differences either in the magnitude recorded by the instrument or in the structure underlying the instrument. Many articles in the *Journal of Cross-Cultural Psychology* fall into this category. For instance, Guida and Ludlow (1989) compared the test anxiety of American and Chilean school children and found that for upper and middle-class subjects, American subjects reported a lower level of test anxiety than Chilean subjects. Two post hoc explanations were then given to explain this finding, none of which was tested in the study.

Two schemes for sampling cultures are employed: systematic and convenience sampling. The subjects can be chosen freely, and as usual, a choice has to be made between matched or random sampling. The procedure will typically amount to the administration of a translated instrument in a new culture. If the instrument has been applied before, the design of the previous study will usually be replicated. Covariates are typically not included in this type of study. The statistical analysis can be based on either level- or structure-oriented techniques. Quite often both aspects are combined; evidence is first provided for the similarity of

psychometric properties (e.g., reliability analysis, factor analysis and target rotations, or analysis of covariance structures), followed by an analysis of variance or *t* test at the item level.

The strength of psychological differences studies is their "open-mindedness" about the presence or absence of cross-cultural differences, a useful strategy to explore cross-cultural differences. When no cross-cultural differences are observed, it is quite likely that neither bias nor intergroup differences exist. The weaknesses of the studies are rather severe. The occurrence of bias is usually not explored. Also, because of the absence of context variables in these studies, the interpretation of the cross-cultural differences observed is not self-evident. It is often difficult to evaluate post hoc interpretations put forward to explain the observed cross-cultural differences. Finally, "fishing" may occur (Cook & Campbell, 1979). It is common that a large number of statistical tests are conducted to test the null hypothesis of no cultural differences. Such multiple testing procedures ("fishing" for significance) can easily lead to the false rejection of the null hypothesis, and hence to incorrect conclusions about the occurrence of cross-cultural differences. Various simple remedies have been proposed in the literature, such as post hoc procedures in analysis of variance or Bonferroni procedures (e.g., Glass & Hopkins, 1984; Hays, 1994). These procedures control for Type I errors when a large number of statistical tests are performed.

External validation studies start from observed cross-cultural differences. These studies aim to identify an appropriate interpretation of the differences. In some cases, external validation is based on previous studies (either generalizability or psychological differences studies) in which cross-cultural differences are reported, while in other cases the observation of cross-cultural differences and external validation are combined in one study. In both cases the choice of cultures, subjects, procedure, and design are straightforward. External validation studies usually involve survey data or secondary data (i.e., data derived from other sources, such as information on national income). External validation studies may be based on various aggregation levels (cf. the section on multilevel modelling). Most frequently reported are the individual level (e.g., when a test of individualism–collectivism is administered and is used as a covariate), an intermediate level (e.g., family and school), and the cultural level (e.g., gross national product, population density). Culture-level data can be derived from various sources such as the Human Relations Area Files (HRAF files; Barry, 1980), other cross-cultural research, and yearbooks of national and international organizations such as OECD, WHO, and UNICEF.

External validation studies are either exploratory in clarifying sources of cross-cultural differences. Analysis of covariance, regression, and causal modelling are the major statistical techniques for studying external validation.

The strength of this approach is its focus on the interpretation of cross-cultural differences, an often neglected issue in cross-cultural psychology. In principle, external validation provides a refutable framework for interpretation. However, the choice of variables and the level of analysis may be arbitrary or meaningless from a psychological point of view. As an example, the distance from a

country's capital to the equator has been found to be a good predictor of various psychological test scores, for example, of cognitive tests. It is obvious that the statistical result does not convey much information about the psychological variables underlying the performance differences.

Examples of this approach can be found in the work of Bond (1991) and Williams and Best (1982). These authors first demonstrated cross-cultural differences (in health measures in Bond's study and in gender stereotypes in Williams and Best's studies). They then related the differences to a wide variety of culture-level measures, such as values, GNP, and per capita expenditure on education and health. The results obtained allow them to interpret the cross-cultural differences observed in terms of these external variables.

Conclusion

The research question or hypothesis, method, and data analysis of cross-cultural studies are closely related. Only properly chosen methods and data analytical procedures will permit an unbiased evaluation of proposed theoretical formulations. In cross-cultural psychology, the interpretation of the meaning of research findings is crucial but evasive. Many interpretations can usually be generated to explain a cross-cultural difference, and it is often difficult to assess their validity. The best approach is to formulate a number of rival hypotheses on an *a priori* basis and design studies that are able to rule out inappropriate explanations. In our opinion, knowledge in cross-cultural psychology accumulates at an unnecessarily slow pace primarily because many cross-cultural researchers rely heavily on post hoc theorizing. This chapter is meant to encourage cross-cultural researchers to place more emphasis on methods and data analysis to improve the effectiveness of their studies. It is also hoped that the chapter will dispel the myth that methodological and statistical sophistication is an obstacle or a distraction in the research enterprise. Quite the contrary, proper methods and data analytical procedures can help clarify conceptual ambiguities, disentangle the influence of confounding variables, and rule out invalid interpretations of cross-cultural differences. Berry (1980) has stated clearly that "Cross-cultural psychology is defined primarily by its *method*" (p. 1, italic in original). Researchers who are committed to cross-cultural research should take methodological issues seriously. This chapter may facilitate cross-cultural researchers to take full advantage of the methodological and statistical procedures available for shaping their work and contributing to the advancement of the field.

References

Allen, M. A., & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA : Brooks/Cole.

Amir, Y., & Sharon, I. (1987). Are social psychological laws cross-culturally valid? *Journal of Cross-Cultural Psychology, 18,* 383–470.

Anastasi, A. (1976). *Psychological testing* (4th ed.). New York: Macmillan.

Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), *Handbook of methods for detecting bias* (pp. 96–116). Baltimore, MD: Johns Hopkins University Press.

Barry, H. (1980). Descriptions and uses of the Human Relations Area Files. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 2, pp. 445–478). Boston: Allyn and Bacon.

Bentler, P. M. (1992). *EQS structural equation program manual.* Los Angeles: BMDP Statistical Software.

Berk, R. A. (Ed.). (1982). *Handbook of methods for detecting item bias.* Baltimore: Johns Hopkins University Press.

Berry, J. W. (1967). Independence and conformity in subsistence-level societies. *Journal of Personality and Social Psychology, 7,* 415–418.

Berry, J. W. (1969). On cross-cultural comparability. *International Journal of Psychology, 4,* 119–128.

Berry, J. W. (1976). *Human ecology and cognitive style: Comparative studies in cultural and psychological adaptation.* Beverly Hills, CA: Sage.

Berry, J. W. (1980). Introduction to methodology. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 1, pp. 1–28). Boston: Allyn and Bacon.

Berry, J. W., & Dasen, P. R. (Eds.) (1974). *Culture and cognition: Readings in cross-cultural psychology.* London: Methuen.

Berry, J. W., Poortinga, Y. H., Segall, M. H., & Dasen, P. R., (1992). *Cross-cultural psychology. Research and applications.* Cambridge: Cambridge University Press.

Berry, J. W., Van de Koppel, J. M. H., Sénéchal, C., Annis, R. C., Bahuchet, S., Cavalli-Sforza, L. L., & Witkin, H. A. (1986). *On the edge of the forest: Cultural adaptation and cognitive development in Central Africa.* Lisse: Swets & Zeitlinger.

Bijnen, E. J., Van der Net, T. Z. J., & Poortinga, Y. H. (1986). On cross-cultural comparative studies with the Eysenck Personality Questionnaire. *Journal of Cross-Cultural Psychology, 17,* 3–16.

Bock, D. (1989). *Multilevel analysis of educational data.* New York: Academic Press.

Bollen, K. J. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. J., & Long, J. S. (Eds.). (1993). *Testing structural equation models.* Newbury Park, CA: Sage.

Bond, M. H. (1991). Chinese values and health: A cross-cultural examination. *Psychology and Health, 5,* 137–152.

Brislin, R. W. (1980). Translation and content analysis of oral and written material. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 1, pp. 389–444). Boston: Allyn and Bacon.

Brislin, R. W., Lonner, W. J., & Thorndike, R. (1973). *Cross-cultural research methods.* New York: Wiley.

Brown, E. D., & Sechrest, L. (1980). Experiments in cross-cultural research. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 2, pp. 297–318). Boston: Allyn and Bacon.

Bryk, A. S. & Raudenbush, W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage.

Buss, D. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12,* 1–49.

Buss, D. M., Abbott, M., Angleitner, A., Asherian, A., Biaggio, A., Blanco-Villasenor, A., Bruchon-Schweitzer, M., Ch'u, H., Czapinski, J., De Raad, B., Ekehammar, B., El Lohamy, N., Fioravanti, M., Georgas, J., Gerde, P., Guttman, R., Hazan, F., Iwawaki, S., Janakiramaiah, N., Khosrokhani, F., Kreitner, S., Lachenicht, L., Lee, M., Liik, M., Little, B., Mika, S., Moadel-Shadid, M., Moane, G., Montero, M., Mundy-Castle, A. C., Niit, T., Nsenduluka, E., Pienkowski, R., Pirttila-Blackman, A-M., Ponce de Leon, J., Rousseau, J., Runco, M. A., Safir, M. P., Samuels, C. Sanitioso, R., Serpell, R., Smid, N., Spencer, C., Tadinac, M., Todorova, E. N., Troland, K., Van den Brande, L., Van Heck, G., Van Langenhove, L., & Yang, K-S. (1990). International preferences in selecting mates. A study of 37 cultures. *Journal of Cross-Cultural Psychology, 21,* 5–47.

Byrne, B. M. (1989). *A primer of LISREL: Basic ap-*

plications and programming for confirmatory factor analytic models. New York: Springer.
Byrne, B. M. (1994). *Structural equation modelling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage.
Campbell, D. T. (1986). Science's social system of validity-enhancing collective believe change and the problems of the social sciences. In D. W. Fiske & R. A. Shweder (Eds.), *Metatheory in social science* (pp. 108–135). Chicago: University of Chicago Press.
Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
Candell, G. L., & Hulin, C. L. (1987). Cross-language and cross-cultural comparisons in scale translations: Independent sources of information about item nonequivalence. *Journal of Cross-Cultural Psychology, 17*, 417–440.
Cattell, R. B. (1940). A culture-free intelligence test, I. *Journal of Educational Psychology, 31*, 176–199.
Cattell, R. B., & Cattell, A. K. S. (1963). *Culture Fair Intelligence Test*. Champaign, IL: Institute for Personality and Ability Testing.
Cheung, F. M. (1989). A review on the clinical applications of the Chinese MMPI. *Psychological Assessment, 3*, 230–237.
Cheung, F. M., Leung, K., Fan, R. M., Song, W. Z., Zhang, J. X., & Chang, J. P. (1996). Development of the Chinese Personality Assessment Inventory (CPAI). *Journal of Cross- Cultural Psychology, 27*, 181–199.
Chinese Culture Connection (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology, 18*, 143–164.
Church, T. A. (1987). Personality research in a non-Western setting: The Philippines. *Psychological Bulletin, 102*, 272–292.
Cicchetti, D. V., Showalter, D., & McCarthy, P. (1990). A computer program for calculating subject-by-subject kappa or weighted kappa coefficients. *Educational and Psychological Measurement, 50*, 153–158.
Clauser, B. E., Mazor, K. M., & Hambleton, R. K. (1994). The effects of score group width on the Mantel-Haenszel procedure. *Journal of Educational Measurement, 31*, 67–78.
Cleary, T. A., & Hilton, T. L. (1968). An investigation of item bias. *Educational and Psychological Measurement, 28*, 61–75.
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.
Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
De Groot, A., Koot, H. M., & Verhulst, F. C. (1994). Cross-cultural generalizability of the Child Behavior Checklist cross-informant syndromes. *Psychological Assessment, 6*, 225–230.
Deregowski, J. B., & Serpell, R. (1971). Performance on a sorting task: A cross-cultural experiment. *International Journal of Psychology, 6*, 273–281.
Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the scholastic aptitude test. *Journal of Educational Measurement, 23*, 355–368.
Draguns, J. (1989). Normal and abnormal behavior in cross-cultural perspective: Specifying the nature of their relationship. *Nebraska Symposium on Motivation, 37*, 235–277. Lincoln, NE: University of Nebraska Press.
Earley, C. (1989). Social loafing and collectivism: A comparison of the United States and the People's Republic of China. *Administrative Science Quarterly, 34*, 565–581.
Eaves, L. J., Eysenck, H. J., & Martin, N. G. (1989). *Genes, culture and personality: An empirical approach*. London: Academic Press.
Ellis, B. B. (1989). Differential item functioning: Implications for test translations. *Journal of Applied Psychology, 74*, 912–921.
Ellis, B. B., & Kimmel, H. D. (1992). Identification of unique cultural response patterns by means of item response theory. *Journal of Applied Psychology, 77*, 177–184.
Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179-197.
Endler, N. S., Lobel, T., Parker, J. D., & Schmitz, P. (1991). Multidimensionality of state and trait anxiety: A cross-cultural study comparing American, Canadian, Israeli and German young adults. *Anxiety Research, 3*, 257–272.
Eysenck, H. J., & Eysenck, S. B. J. (1983). Recent advances in the cross-cultural study of personality. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 2, pp. 41–69). Hillsdale, NJ: Erlbaum.
Feldman, S. S., Rosenthal, D. A., Mont-Reynaud, R., Leung, K., & Lau, S. (1991). Ain't misbehavin': Adolescent values and family environments as correlates of misconduct in Australia, Hong Kong, and the United States. *Journal of Research on Adolescence, 1*, 109–134.
Frijda, N., & Jahoda, G. (1966). On the scope and methods of cross-cultural research. *International Journal of Psychology, 1*, 109–127.
Glass, G. V., & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.
Guida, F. V., & Ludlow, L. H. (1989). A cross-cultural study of test anxiety. *Journal of Cross-Cultural Psychology, 20*, 178–190.
Hakstian, A. R., & Whalen, T. E. (1976). A *k*-sample significance test for independent alpha coefficients. *Psychometrika, 41*, 219–231.
Hambleton, R. K. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment, 9*, 57–68.
Hambleton, R. K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment (Bulletin of the International Test Commission), 10*, 229–244.
Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Dordrecht: Kluwer-Nijhoff.
Harkness, S., & Super, C. M. (1990). Culture and psychopathology. In M. Lewis & S. M. Miller (Eds.), *Handbook of developmental psychopathology. Perspectives in developmental psychology* (pp. 41–52). New York: Plenum Press.
Harman, H. H. (1976). *Modern factor analysis* (3rd rev. ed.). Chicago: University of Chicago Press.
Hays, W. L. (1994). *Statistics* (5th ed.). Orlando, FL: Harcourt Brace & Company.
Hess, R. D., Chang, C. M., & McDevitt, T. M. (1987). Cultural variations in family beliefs about children's performance in mathematics: Comparisons among People's Republic of China, Chinese-American, and Caucasian-American families. *Journal of Educational Psychology, 79*, 179–188.
Ho, D. Y. F. (in press). Filial piety and its psychological consequences. In M. H. Bond (Ed.), *Handbook of Chinese Psychology*. Hong Kong: Oxford University Press.
Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
Hofstede, G. (1983). Dimensions of national cultures in fifty countries and three regions. In J. B. Deregowski, S. Dziurawiec, & R. C. Annis (Eds.), *Expiscations in cross-cultural psychology* (pp. 335–355). Lisse: Swets & Zeitlinger.
Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Erlbaum.
Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.
Hulin, C. L. (1987). Psychometric theory of item and scale translations: Equivalence across languages. *Journal of Cross-Cultural Psychology, 18*, 115–142.
Irvine, S. H. (1979). The place of factor analysis in cross-cultural methodology and its contribution to cognitive theory. In L. Eckensberger, W. Lonner, & Y. H. Poortinga (Eds.), *Cross-cultural contributions to psychology* (pp. 300–341). Lisse: Swets & Zeitlinger.
Irvine, S. H., & Carroll, W. K. (1980). Testing and assessment across cultures. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology* (Vol. 2, pp. 181–244). Boston: Allyn and Bacon.
Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
Jöreskog, K. G., & Sörbom, D. (1993) *LISREL 8*. Chicago: Scientific Software International.

Kaiser, H. F., Hunka, S., & Bianchini, J. (1971). Relating factors between studies based upon different individuals. *Multivariate Behavioral Research, 6,* 409–422.

Kendall, I. M., Verster, M. A., & Von Mollendorf, J. W. (1988). Test performance of blacks in South Africa. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 299–339). Cambridge: Cambridge University Press.

Kraemer, H. C. (1981). Extension of Feldt's approach to testing homogeneity of coefficients of reliability. *Psychometrika, 46,* 41–45.

Lee, S. Y. (1990). Multilevel analysis of structural equation models. *Biometrika, 77,* 763–772.

Leung, K. (1987). Some determinants of reactions to procedural models for conflict resolution. *Journal of Personality and Social Psychology, 53,* 898–908.

Leung, K. (1989). Cross-cultural differences: Individual-level vs. culture-level analysis. *International Journal of Psychology, 24,* 703-719.

Leung, K., Au, Y., Fernandez-Dols, J. M., & Iwawaki, S. (1992). Preference for methods of conflict processing in two collectivist cultures. *International Journal of Psychology, 27,* 195-209.

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

Leung, K., & Drasgow, F. (1985). Relation between self-esteem and delinquent behavior in three ethnic groups: An application of item response theory. *Journal of Cross-Cultural Psychology, 17,* 151–167.

Leung, K., & Zhang, J. X. (1995). Systemic consideration: Factors facilitating and impeding the development of psychology in developing countries. *International Journal of Psychology, 30,* 693–706.

Little, T. D., Oettingen, G., & Baltes, P. B. (1995). *The revised Control, Agency, and Means- Ends Interview (CAMI).* Berlin: Max-Planck-Institut für Bildungsforschung.

Little, T. D., Oettingen, G., Stetsenko, A., & Baltes, P. B. (1995). Children's action-control beliefs and school performance: How do American children compare with German and Russian children? *Journal of Personality and Social Psychology, 69,* 686–700.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68,* 304–305.

Lord, F. M. (1977). A study of item bias, using Item Characteristic Curve Theory. In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural psychology* (pp. 19–29). Lisse: Swets & Zeitlinger.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Marsh, H. W., & Byrne, B. M. (1993). Confirmatory factor analysis of multigroup-multimethod self-concept data: Between-group and within-group invariance constraints. *Multivariate Behavioral Research, 28,* 313–349.

McCrae, R. R., & Costa, P. T. (1985). Updating Norman's "adequacy taxonomy": Intelligence and personality dimensions in natural language and in questionnaires. *Journal of Personality and Social Psychology, 49,* 710–721.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality, 60,* 175–215.

McCauley, C. D., & Mendoza, J. (1985). A simulation study of item bias using a two-parameter item response model. *Applied Psychological Measurement, 9,* 389–400.

McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics, 7,* 105–118.

Ostroff, C. (1993). Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology, 78,* 569–582.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart, & Winston.

Peterson M. F., Smith, P. B., Akande, A., Ayestaran, S., Bochner, S., Callan, V., Cho, N. G., Jesuino, J. C., D'Amorim, M., Francois, P., Hofmann, K., Koopman, P. L., Leung, K., Lim, T. K., Mortazavi, S., Munene, J., Radford, M., Ropo, A., Savage, G., Setiadi, B., Sinha, T. N., Sorenson, R., & Viedge, C. (1995). Role conflict, ambiguity, and overload: A 21-nation study. *Academy of Management Journal, 38,* 429–452.

Poon, W. Y., Chan, W., Lee, S. Y., & Leung, K. (1993). Preliminary analysis of multiple group structural equation modelling via cluster analysis. *Proceedings of the American Statistical Association 1993 Convention, Social Statistics Section,* 368–373.

Poortinga, Y. H. (1971). Cross-cultural comparison of maximum performance tests: Some methodological aspects and some experiments with simple auditory and visual stimuli. *Psychologia Africana,* Monograph Supplement, No. 6.

Poortinga, Y. H. (1989). Equivalence of cross-cultural data: An overview of basic issues. *International Journal of Psychology, 24,* 737–756.

Poortinga, Y. H., & Malpass, R. S. (1986). Making inferences from cross-cultural data. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural psychology* (pp. 17–46). Beverly Hills, CA: Sage.

Poortinga, Y. H., & Van de Vijver, F. J. R. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology, 18,* 259–282.

Poortinga, Y.H., Van de Vijver, F. J. R., Joe, R. C., & Van de Koppel, J. M. H. (1987). Peeling the onion called culture: A synopsis. In C. Kagitcibasi (Ed.), *Growth and progress in cross-cultural psychology* (pp. 22–34). Lisse: Swets & Zeitlinger.

Poortinga, Y. H., & Van der Flier, H. (1988). The meaning of item bias in ability tests. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 166–183). Cambridge: Cambridge University Press.

Reuning, H., & Wortley, W. (1973). Psychological studies of the Bushmen. *Psychologia Africana,* Monograph Supplement, 7.

Roberts, J., & Sutton-Smith, B. (1962). Child training and game involvement. *Ethology, 1,* 166–185.

Russell, J. A., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology, 57,* 848–856.

Sachs, J. (1992). Covariance structure analysis of a test of moral orientation and moral judgment. *Educational and Psychological Measurement, 52,* 825–833.

Scheuneman, J. D. (1987). An experimental, exploratory study of causes of bias in test items. *Journal of Educational Measurement, 24,* 97–118.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). Orlando, FL: Academic Press.

Schwartz, S. H. (1994). Studying human values. In A. Bouvy, F. J. R. Van de Vijver, P. Boski, & P. Schmitz (Eds.), *Journeys into cross-cultural psychology* (pp. 239–254). Lisse: Swets & Zeitlinger.

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception.* Indianapolis, IN: Bobbs-Merrill.

Serpell, R. (1979). How specific are perceptual skills? *British Journal of Psychology, 70,* 365-380.

Serpell, R. (1993). *The significance of schooling. Life-journeys in an African society.* Cambridge: Cambridge University Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

Smith, P. B., & Peterson, M. F. (1988). *Leadership, organizations and culture.* Beverly Hills, CA: Sage.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology, 49,* 607–627.

Stricker, L. J. (1982). Identifying test items that perform differentially in population subgroups: A partial correlation index. *Applied Psychological Measurement, 6,* 261–273.

Super, C. M. (1981). Behavior development in infancy. In R. H. Munroe, R. L. Munroe, & B. B. Whiting (Eds.), *Handbook of cross-cultural human development* (pp. 181–270). New York: Garland SPTM Press.

Super, C. M. (1983). Cultural variation in the meaning and uses of children's "intelligence." In J. B. Deregowski, S. Dziurawiec, & R. C. Annis (Eds.), *Expiscations in cross-cultural psychology* (pp. 199–212). Lisse: Swets & Zeitlinger.

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington, DC: Department of the Army.

Tylor, E. B. (1871). *Primitive culture* (2 vols.). London: Murray.

Van de Vijver, F. J. R. (1988). Systematizing item content in test design. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 291–307). New York: Plenum.

Van de Vijver, F. J. R. (1994). Item bias: Where psychology and methodology meet. In A. Bouvy, F. J. R. Van de Vijver, P. Boski, & P. Schmitz (Eds.), *Journeys into cross-cultural psychology* (pp. 111–126). Lisse: Swets & Zeitlinger.

Van de Vijver, F. J. R., Daal, M., & Van Zonneveld, R. (1986). The trainability of abstract reasoning: A cross-cultural comparison. *International Journal of Psychology, 21,* 589–615.

Van de Vijver, F. J. R., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized version of the General Aptitude Test Battery. *Journal of Applied Psychology, 79,* 852–859.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1982). Cross-cultural generalization and universality. *Journal of Cross-Cultural Psychology, 13,* 387–408.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1991). Testing across cultures. In R. K. Hambleton & J. Zaal (Eds.), *Advances in educational and psychological testing* (pp. 277–308). Dordrecht: Kluwer.

Van de Vijver, F. J. R. & Poortinga, Y. H. (1992). Testing in culturally heterogeneous populations: When are cultural loadings undesirable? *European Journal of Psychological Assessment, 8,* 17–24.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1994). Methodological issues in cross-cultural studies on parental rearing behavior and psychopathology. In C. Perris, W. A. Arrindell, M. Eisemann (Eds.), *Parental rearing and psychopathology* (pp. 173–197). Chicester: Wiley.

Van den Wollenberg, A. L. (1988). Testing a latent trait model. In R. Langeheine & J. Rost (Eds.), *Latent trait and latent class models* (pp. 31–50). New York: Plenum.

Van Haaften, E. H., & Van de Vijver, F. J. R. (in press). Psychological consequences of environmental degradation, *Journal of Health Psychology.*

Vandenberg, R. J., & Self, R. M. (1993). Assessing newcomers' changing commitments to the organization during the first 6 months of work. *Journal of Applied Psychology, 78,* 557–568.

Vernon, P. E. (1969). *Intelligence and cultural environment.* London: Methuen.

Vernon, P. E. (1979). *Intelligence: Heredity and environment.* San Francisco: Freeman.

Watkins, D. (1989). The role of confirmatory factor analysis in cross-cultural research. *International Journal of Psychology, 24,* 685–702.

Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of cultural anthropology* (pp. 398–419). New York: American Museum of Natural History.

Whiting, B. B. (1976). The problem of the packaged variable. In K. Riegel & J. Meacham (Eds.), *The developing individual in a changing world* (Vol. 1, pp. 303–309). The Hague: Mouton.

Williams, J. E., & Best, D. L. (1982). *Measuring sex stereotypes: A thirty-nation study.* Beverly Hills, CA: Sage.

Wilss, W. (1982). *The science of translation: Problems and methods.* Tuebingen: Narr.

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw Hill.

Yang, K. S., & Bond, M. H. (1990). Exploring implicit personality theories with indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology, 58,* 1087–1095.

Zegers, F. E., & Ten Berge, J. M. F. (1985). A family of association coefficients for metric scales. *Psychometrika, 50,* 17–24.

8

CULTURE AS PROCESS: EMPIRICAL METHODS FOR CULTURAL PSYCHOLOGY

PATRICIA M. GREENFIELD[1]
University of California, Los Angeles
United States