Approaches to equivalence in cross-cultural and cross-national survey research

Johnson, Timothy P.

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research

TIMOTHY P. JOHNSON

In cross-cultural (and cross-national) survey research, the equivalence of survey questions rivals the importance of their reliability and validity. This paper presents a review of the multiple dimensions of equivalence that must be addressed when conducting comparative survey research. Available methodologies for establishing one or more forms of equivalence are also identified and the strengths and limitations of each approach are examined. It is concluded that multiple methodologies must be implemented in order to insure the cross-cultural equivalence of survey measures.

1. Introduction

In perhaps no other subfield of social science research are issues of methodology and measurement as open to challenge and criticism as when they are applied in cross-cultural and cross-national settings. Indeed, the available protocols for conducting cross-cultural and cross-national survey research (to be subsequently referred to as cross-cultural survey research) would appear to be seriously underdeveloped in comparison to the methodologies available for the conduct of monocultural surveys. A major source of the criticism directed at cross-cultural survey research, in fact, has been the uncritical adaptation of the highly successful techniques developed for monocultural surveys. The simple application of this technology in cross-cultural settings usually and unfortunately makes gross assumptions regarding the equivalence of concepts and measurement. Although this problem is recognized by most practitioners, and many have made serious

attempts to address it, there is currently little consensus regarding how best to establish cross-cultural equivalence when conducting social surveys.

One possible explanation for this absence of methodological consensus, given the sheer quantity of cross-cultural surveys that have been conducted over many decades, is perhaps an even more fundamental lack of agreement regarding the notion of equivalence. As we shall shortly see, researchers concerned with cross-cultural inquiries have conceptualized and cataloged equivalence in numerous ways. It would seem obvious that differing views of what equivalence is would almost certainly lead to variability in the procedures proposed for investigating or establishing it. The purpose of this paper is to present an investigation of these closely-related problems. Specifically, it will review: (1) the concept of equivalence as it has been applied to cross-cultural survey research; and (2) available methodologies that have been proposed and/or previously implemented for the purpose of assessing or implementing one or more forms of cross-cultural equivalence when conducting social surveys.

2. Notions of "Equivalence"

Common sense definitions of the term "equivalent" include: "equal in force, amount, or value;" and "corresponding or virtually identical especially in effect or function" (Webster's Seventh New Collegiate Dictionary, 1965). Perhaps in no field of inquiry, though, has this seemingly elementary concept been assigned as many alternative meanings and disaggregated into as many components as in the field of cross-cultural research. Table 1 presents the results of an investigation of the types of "equivalence" that have been discussed or mentioned in the available literature on cross-cultural research. This review included work representing the disciplines of anthropology, business, communication, demography, economics, market research, political science, psychiatry, psychology, sociology, as well as other professions, and covered work reported over the past 35 years.

As Table 1 indicates, more than 50 specific terms have been used to discuss varieties of equivalence. Some of these have not been well defined. As might be expected, there is also considerable overlap, and many of these alternative labels probably represent "equivalent" concepts (see below). Two of the terms used in this table, *cross-cultural equivalence* (Hui and Triandis, 1985) and *cultural equivalence* (Devins et al., 1997) appear to have been used in a generic sense, referring collectively to all forms of equivalence. They will be used in a similar manner in this review. In addition, although it was not my intention in conducting this review to contribute to the plethora of equivalence labels inhabiting the literature, for purposes of parsimony, the remaining forms of equivalence listed in Table 1 can be subsumed under what can be defined as two fundamental domains of cross-cultural equivalence: interpretive and procedural. These two general domains will be examined in turn.

2.1 Table 1: Types of Equivalence Referenced in the Literature

1. ***Calibration Equivalence*** - Mullen (1995)
2. ***Complete Equivalence*** - Verba et al. (1978)
3. ***Conceptual Equivalence*** - Adams-Esquivel (1991); Elder (1976); Eyton and Neuwirth (1984); Flaherty et al. (1988); Green and White (1976); Hines (1993); Hui and Triandis (1985); Kohn and Slomczynski, 1990; Miller et al. (1981); Mitchell (1973); Narula (1990); Okazaki and Sue (1995); Sears (1961); Sechrest et al. (1972); Singh (1995); Straus (1969); Warwick and Osherson (1973)
4. ***Construct Equivalence*** - Singh (1995); Van de Vijver and Leung (forthcoming)
5. ***Construct Operationalization Equivalence*** - Hui and Triandis (1983)
6. ***Content Equivalence*** - Flaherty et al. (1988)
7. ***Contextual Equivalence*** - Elder (1973)
8. ***Credible Equivalence*** - Teune (1990)
9. ***Criterion Equivalence*** - Flaherty et al. (1988)
10. ***Cross-cultural Equivalence*** - Devins et al. (1997); Hui and Triandis (1985); Hui

et al. (1983)

11. ***Cultural Equivalence*** - Devins et al. (1997)
12. ***Definitional Equivalence*** - Eyton and Neuwirth (1984)
13. ***Direct Equivalence*** - Frey (1970)
14. ***Exact Equivalence*** - Verba et al. (1978)
15. ***Experiential Equivalence*** - Sechrest et al. (1972)
16. ***Factor Equivalence*** - Dressler et al. (1991)
17. ***Factorial Equivalence*** - Singh (1995)
18. ***Formal Equivalence*** - Frey (1970); Marsh (1967); Miller et al. (1985); Mohler et al. (1996)
19. ***Functional Equivalence*** - Alwin et al (1994); Allerbeck (1977); Berry (1969); Braun and Scott (1996); Czudnowski (1976); Frey (1970); Frijda and Jahoda (1966); Green and White (1976); Hui and Triandis (1983; 1985); Marsh (1967);Mitchell (1973); NieBen (1982); Pareek and Rao (1980); Peschar (1982); Scheuch, (1993); Sekaran (1983); Singh (1995); Teune (1990); Van de Vijver and Poortinga (1982); Verba (1969); Verba et al. (1978)
20. ***Grammatical-Syntactical Equivalence*** - Sechrest et al. (1972)
21. ***Indicator Equivalence*** - Kuechler (1987)
22. ***Idiomatic Equivalence*** - Sechrest et al. (1972)
23. ***Instrument Equivalence*** - Frey (1970); Green and White (1976); Singh (1995)
24. ***Item Equivalence*** - Borg (1996); Hui and Triandis (1983; 1985); Mohler et al. (1996)
25. ***Lexical Equivalence*** - Blumer and Warwick (1993); Deutscher (1973); Elder (1973); Warwick and Osherson (1973)
26. ***Linguistic Equivalence*** - Berry et al. (1992); Ellis et al. (1989); Hines (1993); Hulin (1987); Iyengar (1993); Kohn and Slomczynski, 1990; Okazaki and Sue (1995); Prince and Mombour (1967); Sechrest et al. (1972); Warwick and Osherson (1973)

27. ***Literal Equivalence*** - Frijda and Jahoda (1966); Mohler et al (1996)
28. ***Meaning Equivalence*** - Prince and Mombour (1967)
29. ***Measurement Equivalence*** - de Vera (1985); Drasgow and Kanfer (1985); Dressler et al. (1991); Ellis et al. (1989); Green and White (1976); Hui et al. (1983); Iyengar (1993); Leung and Drasgow (1986); Mullen (1995); Poortinga (1989); Singh (1995); Straus (1969)
30. ***Measurement Unit Equivalence*** - Van de Vijver and Leung (1996)
31. ***Metaphorical Equivalence*** - Dunnigan et al. (1993)
32. ***Metric Equivalence*** - Hui and Triandis (1983); Leung and Bond (1989); Mullen (1995); Okazaki and Sue (1995); Straus (1969); Van de Vijver and Leung (1996); Van de Vijver and Poortinga (1982)
33. ***Motivational Equivalence*** - Triandis (1972)
34. ***Operational Equivalence*** - Mohler et al (1996); Narula (1990); Prince and Mombour (1967)
35. ***Psychological Equivalence*** - Eckensberger (1973)
36. ***Psychometric Equivalence*** - Devin et al. (1997); Ellis et al. (1989); Hulin (1987); Van de Vijver and Poortinga (1982)
37. ***Relational Equivalence*** - Ellis et al. (1989)
38. ***Relative Equivalence*** - Frey (1970)
39. ***Response Equivalence*** - Anderson (1967); Frey (1970); Sekaran (1983)
40. ***Scalar Equivalence*** - Hui and Triandis (1983; 1985); Mullen (1995); Van de Vijver and Leung (1996); Van de Vijver and Poortinga (1982)
41. ***Semantic Equivalence*** - Flaherty et al (1988); Kleinman (1987)
42. ***Situational Equivalence*** - Anderson (1967)
43. ***Stimulus Equivalence*** - Anderson (1967); Verba et al. (1978)
44. ***Structural Equivalence*** - Van de Vijver and Leung (1996); Watkins(1989)
45. ***Substantive Equivalence*** - Czudnowski (1976)
46. ***Syntactic Equivalence*** - Kohn and Slomczynski (1990)

47. ***Technical Equivalence*** - Flaherty et al (1988)
48. ***Text Equivalence*** - Alwin et al (1994)
49. ***Theoretical Equivalence*** - Teune (1977; 1990)
50. ***Translation Equivalence*** - Anderson (1967); Berry et al. (1992); Candell and Hulin (1987); Hui and Triandis (1983); Hulin (1987); Mullen (1995)
51. ***Verbal Equivalence*** - Adams-Esquivel (1991)
52. ***Vocabulary Equivalence*** - Sechrest et al. (1972)

2.2 Interpretive Equivalence

Several types of equivalence that have been discussed in the literature are primarily concerned with similarities in how abstract, or latent, concepts are interpreted across cultures. As such, these types are very similar in their emphasis on equivalence of meaning, and will consequently be classified as subtypes of "interpretive" equivalence. One of the more commonly cited forms is *conceptual equivalence*, which Hui and Triandis (1985) would apply to constructs that can be meaningfully discussed within each of the cultures of interest. They identify conceptual equivalence as a necessary condition for making cross-cultural comparisons. Similarly, Okazaki and Sue (1995) associate conceptual equivalence with the degree to which a particular concept has identical meaning within two or more cultural groups.

An emphasis on concordance of meaning also appears to be the central requirement for *functional equivalence*. In discussing this form, Van de Vijver and Poortinga (1982) state that "concepts with functional equivalence are universal in a qualitative, although not necessarily a quantitative sense." Pareek and Rao (1980) also emphasize the commonality of meaning across cultures when discussing functional equivalence, suggesting that it "exists when the behavior in question has developed in response to a problem shared by two or more social/cultural groups, even though the behavior in one society may be superficially quite different from the behavior in another society." Additionally, Singh (1995) argues that functional equivalence exists to the degree that the

concept serves similar functions within each society being investigated. *Definitional equivalence*, as discussed by Eyton and Neuwirth (1984), would appear to have a similar meaning.

Other forms of equivalence that have been discussed in the literature also appear to be primarily concerned with meaning. One of these is *semantic equivalence*, a concept which Flaherty et al. (1988) would apply to survey items that exhibit identical meaning across two or more cultures after translation. Similarly, Prince and Mombour (1967) define questionnaires that have successfully retained their original meaning after translation as having *linguistic equivalence.* Iyengar (1993) uses the same label to describe questionnaires that have validity across two or more languages. *Translation equivalence* (Hui and Triandis, 1983), *meaning equivalence* (Prince and Mombour, 1967), and *contextual equivalence* (Elder, 1973) would also appear to be concerned with similarity of construct interpretation across groups. Similarly, Sechrest et al. (1972) discuss *idiomatic equivalence*, which refers to the equivalence or inequivalence of idiomatic expressions used in survey items across cultural groups. Finally, three other terms that have been put forth by researchers, *experiential equivalence* (Sechrest et al., 1972), *theoretical equivalence* (Teune, 1977) and *substantive equivalence* (Czudnowski, 1976), are concerned with the cross-group similarity of the social processes being investigated.

2.3 Procedural Equivalence

A second form of equivalence that has been discussed at varying levels of detail in the literature is concerned with the measures and procedures used to make cross-cultural comparisons. For purposes of this review, these concepts will be defined as subtypes of "procedural" equivalence. One of these includes forms which focus on cross-cultural consistency of measurement. Among these are *exact equivalence* (Verba et al., 1978), *lexical equivalence (*Warwick and Osherson, 1973), *literal equivalence* (Frijda and

Jahoda, 1966), *verbal equivalence* (Adams-Esquivel, 1991), *vocabulary equivalence* (Sechrest et al., 1972), and perhaps also *indicator equivalence* (Kuechler, 1987) *stimulus equivalence* (Anderson, 1967) and *text equivalence* (Alwin et al., 1994), each of which suggests or implies a strict similarity of question wording across language groups. Related forms of equivalence include *formal equivalence* (Frey, 1970), *instrument equivalence* (Singh, 1995), *item equivalence* (Hui and Triandis, 1985), *measurement equivalence* (Leung and Drasgow, 1986), *psychometric equivalence* (Hulin, 1987), *syntactic equivalence* (Kohn and Slomczynski, 1990), and *grammatical-syntactical equivalence* (Sechrest et al., 1972), each of which emphasize the applicability of mechanically identical procedures across groups. Experienced researchers recognize both the pitfalls of uncritically assuming these forms of equivalence and the difficulties of formally demonstrating their presence. These concepts often represent what Berry (1969) has referred to as an "imposed etic" process, in that survey instruments initially designed for one culture are subsequently adapted in a strict technical sense for use with other cultural groups.

Another set of procedural equivalence concepts are concerned with varying levels of psychometric comparability among cross-cultural samples. *Metric equivalence*, for example, is thought to exist when survey questions exhibit similar statistical properties when measured across varying cultural groups (Hui and Triandis, 1983; Okazaki and Sue, 1995; Straus, 1969; Van de Vijver and Leung, 1996). Even more precisely, *measurement unit equivalence* exists when a measurement scale is identical across groups, but there is no common origin (Van de Vijver and Leung, 1996). When measures also have a common origin across groups, they are considered to have *scalar equivalence* (Hui and Triandis, 1983; 1985; Van de Vijver and Poortinga, 1982; Van de Vijver and Leung, 1996) or *calibration equivalence* (Mullen, 1995). *Structural equivalence* assesses the degree to which survey data collected across cultures produce equal data structures, such as what might be observed using factor analysis and similar procedures (Van de Vijver and Leung, 1996). *Factor equivalece* is also concerned with similarity of data structures,

but only to the degree that equal numbers of factors are identified across cultures via factor analysis. *Factorial equivalence* is concerned with the degree to which factor loadings are similar across cultural groups (Singh, 1995). Finally, *measurement equivalence*, as defined by Singh (1995; although see competing definitions provided by Leung and Drasgow (1986) and Straus, 1969), represents instances in which both factor loadings and error variances are identical across groups. A strict burden of equivalence indeed!

Frey (1970) discusses procedural equivalence from the perspective of the cross-cultural equating of measures. Specifically, he discusses *direct equivalence* as existing when measures can be directly compared across cultural groups without reference to culture-specific criteria. In contrast, *relative equivalence* exists when measures collected across two or more cultures must be standardized in reference to some other norm or criteria before they can be compared. For example, annual income can be reasonably compared across nations, but only after being standardized to one metric.

Another cluster of concepts share a concern with the cross-cultural validation of survey items and/or survey scales. Hui and Triandis (1983), for example, discuss *construct operationalization equivalence* as being a form of construct validity. A measure can be identified as having this type of equivalence to the degree that it exhibits a consistent theoretically-derived pattern of relationships with other variables across the cultural groups being examined. *Construct equivalence* (Singh, 1995) and *relational equivalence* (Ellis et al., 1989) would appear to have much the same meaning. *Criterion equivalence*, in contrast, is concerned with the degree to which a variable is consistently associated with other measures of the same construct across cultural groups (Flaherty et al., 1988). Flaherty et al. (1988) also discuss *content equivalence*, which they identify as being the extent to which the items in a measurement scale adequately represent the theoretical domain of interest within each culture being examined. Eckensberger (1973) assigns a very similar meaning to the term *psychological equivalence*. One additional form is *response equivalence*, which Frey (1970) defines as the degree to which responses

obtained from bilingual persons are similar when expressed in two or more different languages.

Both *situational* (Anderson, 1967) and *technical* (Flaherty et al., 1988) *equivalence* are concerned with the conditions under which surveys are administered. Of primary concern here is that the method of data collection used within each culture produces a similar stimulus. *Motivational equivalence* (Triandis, 1972) reflects an interest in assessing the degree to which respondents from varying cultures have similar motivations for their responses.

Another form of procedural equivalence has been referred to as *operational equivalence*. Although its use by Prince and Mombour (1967) is somewhat vague, Mohler et al. (1996) refer to measures as having operational equivalence if "one can be substituted for the other with no detectable change in statistical analyses."

Finally, without distinguishing between interpretational and procedureal forms of equivalence, Verba et al. (1978) refer to *complete equivalence* as a hypothetical achievement that will never be attainable in practice. In contrast, Teune's (1990) discussion of *credible equivalence* implies that some minimum level of either interpretational or procedural similarity may need to be demonstrated in practice before cross-cultural comparisons can be made.

How are these various types of equivalence established within the context of cross-cultural survey research? Just as there are multiple forms of equivalence with which researchers must be concerned, there are numerous methodological approaches that may be useful for addressing them. It is this issue to which our attention is next directed.

3. Available Methods for Establishing Equivalence

In reviewing available research methodologies for assessing cross-cultural equivalence in survey measurement, it may be useful to utilize the "etic-emic" conceptual model (Berry,

1969; Triandis, 1972) from anthropology and psychology. According to this framework, concepts, ideas and behaviors represented by survey questions can be classified as universal or "etic" to the degree that they are universal, or understood in a consistent manner across cultural and national boundaries (i.e., to the extent that they have interpretive equivalence). In contrast, some ideas and concepts are considered "emic" if they have meaning only to one or a few cultural groups, that is, if they are culture-specific or nation-specific.

Interpretive equivalence can never be established for emic phenomena because they do not have shared meaning across cultures. Some forms of procedural equivalence, ironically, can be obtained for emic phenomena. Survey instruments, for example, may impose identical wording on survey questions that are to be used across cultural groups, even if the concepts represented by those questions are emic to a single group. This, however, would be most appropriately referred to as a pseudoetic application of an emic construct. As mentioned earlier, Berry (1969) would refer to such an application as an "imposed etic" practice. This terminology will be useful throughout the remainder of this review.

The techniques which have been applied to problems of cross-cultural equivalence in survey research have been organized around four specific phases of survey research projects: question development, questionnaire pretesting, data collection, and data analysis (see Table 2). It should be noted that the discussion of each technique is intended to serve as a brief overview and not as a comprehensive presentation. References are provided for readers interested in obtaining additional information regarding any of these approaches.

Table 2: Available Methods for Addressing Equivalence in Cross-Cultural Survey Research

A. ***Question Development Phase***

(1). Expert consultation/collaboration

(2). Ethnographic and other qualitative approaches

(3). "Good" question wording practices

(4). "Good" translation practices

(5). Facet analysis

B. ***Questionnaire Pretesting Phase***

(6). Cognitive interviews/structured probes

(7). Measuring response category intensity

(8). Comparative behavior coding

(9). Compare alternative data collection modes

C. ***Data Collection Phase***

(10). Use multiple indicators

(11). Use both emic and etic questions

(12). Respondent/interviewer matching

D. ***Data Analysis Phase***

(13). Item analysis

(14). Item response theory

(15). Generalizability theory

(16). Confirmatory factor analysis

(17). Multidimensional scaling

(18). Applying statistical controls

(19). Identity-equivalence method

3.1 Question Development Phase

Perhaps the most intuitive method for improving the interpretive equivalence of survey questionnaires is the active participation of experts representative of each culture to be studied. This participation may take a number of forms. Two of the primary ones have been *expert consultation* and *expert collaboration*. Examples of expert consultation include: (a) Straus' (1969) proposal to employ cultural experts as judges for evaluating the appropriateness of specific survey items within their culture; and (b) Henderson et al.'s (1992) recommendation that members of each culture being examined be consulted in order to assure that topics of relevance to them are considered. Berry et al., (1992), Elder (1976) and Okazaki and Sue (1995) have each suggested a similar approach. Flaherty et al. (1988) have made more detailed recommendations for expert consultation, suggesting that such teams should include both content specialists and social scientists from each culture. Such teams would be asked to review the appropriateness of instrument content and data collection methods, and to identify other culture-specific considerations. A team or committee approach to questionnaire translation has also been recommended by several researchers (Adams-Esquivel, 1991; Brislin, 1986; Jones and Kay, 1992; Werner and Campbell, 1970). Although clearly very helpful, consultation is not the same as collaboration and may sometimes carry with it some of the less desirable connotations of "hired-hand" research, such as lack of commitment and status inconsistencies.

Others have emphasized more formal integration of cultural representatives as full research collaborators. Frey (1970), for example, has written that "the basic procedure is to assemble a research group possessing deep familiarity with the nations to be studied and with existing research techniques. This group must agree on the objectives of the research and reach a mutual understanding of its major concepts and hypotheses." More recently, Van de Vijver and Hambleton (1996) have stated that "successful avoidance of ethnocentric tendencies in instruments may require a multicultural, multilingual team

with an expertise in the construct under study." Brislin (1986), Johnson et al. (1996a), Kuechler (1987) and Triandis (1972) have made similar recommendations In the United States, the active collaboration of representatives from all participating cultural groups is now often a requirement for the receipt of research funding from federal agencies. The advantages of this approach for assessing and contributing to interpretive equivalence are clear and there appear to be few disadvantages. However, most of the recommendations cited above tend to emphasize collaboration only during the early hypothesis development and questionnaire design phases of research efforts. At the risk of stating the obvious, it is also important to recognize that collaboration should continue throughout all stages of the research process.

Ethnographic and other qualitative approaches have also been recommended as methods for developing interpretively equivalent survey measures. Marin and Marin (1991), for example, suggest cultural immersion, contact with informants, and familiarity with the available literature as appropriate means of improving cultural awareness prior to study design and question development. Word (1992) has also indicated that, prior to constructing survey instruments, ethnographic research may be useful for achieving a more in-depth understanding of the cognitive processes used by persons in different cultures. While these procedures offer obvious advantages, many researchers unfortunately find them less attractive because they are often time-consuming (Ferketich, Phillips and Verran, 1993). For those without the resources to conduct their own ethnographic inquiries, useful information may nonetheless be obtained from the Human Relations Area Files (HRAF), a large data base that maintains information regarding hundreds of unique social and cultural groups (Barry, 1980; Marsh, 1967).

There are also other less intensive qualitative strategies that may be employed during the development of survey questionnaires. One such approach is the antecedent-consequent method described by Triandis (1977). The method is both simple and powerful. Respondents representing the cultures of interest are asked to contribute phrases to a

series of incomplete sentences in order to complete them. By doing so, they can provide researchers with important insights into cross-cultural similarities and differences in perceptions of both the causes and consequences of various phenomena. Another approach is to ask respondents to perform card sorting tasks. These exercises can provide comparative information regarding how respondents organize and manipulate domains of content information. Johnson et al. (forthcoming), for example, have successfully employed this technique to investigate the social identities of multiracial individuals. Focus groups, of course, are a well-known qualitative technique that can provide additional insights when formulating survey questions for use in cross-cultural surveys (Harari and Beaty, 1990). Other qualitative approaches are discussed by Hines (1993).

Adherence to *"good" question wording practices* is another method that focuses primarily on procedural equivalence. Although there is no consensus on what those best practices might be, Brislin (1973; 1986) has over several decades refined a set of general principles that have received considerable attention. In brief, these include the following (Brislin, 1986):

(1). Use short, simple sentences of less than sixteen words;
(2). Employ the active rather than the passive voice;
(3). Repeat nouns instead of using pronouns;
(4). Avoid metaphors and colloquialisms;
(5). Avoid the subjunctive;
(6). Add sentences to provide context for key ideas;
(7). Avoid adverbs and prepositions telling "where" or "when;"
(8). Avoid possessive forms where possible;
(9). Use specific rather than general terms;
(10). Avoid words indicating vagueness regarding some event or thing;
(11). Use wording that will be familiar to translators; and
(12). Avoid sentences with two different verbs if the verbs suggest two different actions.

Bernard (1988) also provides a basic set of recommendations for the development of survey questions that are to be used cross-culturally.

Suggestions for "good" wording practices that will contribute to successful question translation have also been offered by several other researchers. Scheuch (1993), for example, posits that more abstract concepts have a greater likelihood of producing differences in meaning across languages and should therefore by avoided when possible. Prince and Mombour (1967) warn that "if there is a discrepancy in the frequency of usage of a word in two cultures, the words do not have meaning equivalence for survey purposes" and should also be avoided. In addition, it has been suggested by Warwick and Osherson (1973) that "one of the most effective aids to linguistic equivalence is a research problem that is salient to the cultures involved." The more relevant a concept is to everyday existence within a culture, they posit, the fewer the difficulties of language and translation that will be experienced. McKay et al. (1996) suggest the avoidance of slang terms. They also suggest avoiding modifiers and providing examples designed to increase comprehension, as these may also contribute to cross-cultural differences in interpretation.

Cultural differences in response styles are also a challenge to interpretive equivalence. For example, the well known "courtesy bias" found in many societies (Jones, 1963) suggests that questions that might invite obviously socially desirable responses should be avoided wherever possible. To further combat this problem, Mitchell (1973) has recommended that "moral" words be avoided when preparing survey questions, as they are also likely to encourage socially desirable responses. Inkeles and Smith (1974) suggest that "agree-disagree" response formats be avoided for the same reasons.

Smith (1988) provides several suggestions for improving the equivalence of the response scales used in cross-cultural studies. One of these is to consider the use of numerical scales, which he argues can "reduce problems by providing a universally understood set

of categories that have precise and similar meanings," and avoid the use of vague quantifiers, which are more likely to exhibit cross-cultural differences in interpretation. He acknowledges that this approach is also less than perfect in that numeric scales are often more complex than the simple Likert-type scales they are designed to replace, and that different cultures may vary in the ways they manipulate numeric information. Another approach suggested by Smith (1988) is the use of simple dichotomous response options, which may be less susceptible to misunderstanding than traditional ordinal response scales. Smith (1997) also provides useful recommendations regarding the use of various response options across cultures. For example, he indicates that symmetrical, bipolar scales with a clear middle point will likely be most successful in cross-cultural studies.

Collectively, these recommendations for "good" question wording practices can be expected in many instances to contribute to the interpretive equivalence of survey questions. These approaches, however, do not necessarily rule out equivalence threats associated with cross-cultural differences in the fundamental understanding of the concepts, ideas and/or behaviors being assessed. The emphasis of this approach to similarity of question wording, even "good" question wording, will always insure some degree of procedural equivalence at the risk of failing to achieve interpretive equivalence. Survey researchers will need to recognize that there are likely to be many etic concepts that can nonetheless not be assessed using identical survey questions across any random pair of cultures. In recognition of this, some have advocated the use of open-ended questions as a method of verifying equivalence of meaning across cultures (Verba et al., 1978).

Over the past several decades, effort has also been invested in the development of *"good" translation practices* for survey questionnaires. It has been clear for some time that a simple, unidirectional translation of a survey instrument from a source language into one or more target languages is an unacceptable procedure. A commonly referenced improvement is the back-translation model (Brislin, 1970; 1976; 1986). Although there

are countless variations (see for example: Anderson, 1967; Frey, 1970; Marin and Marin, 1991), the basic procedure calls for a bilingual person to translate a source questionnaire into a target language. A second bilingual person is then asked to translate this version back into the source language without knowledge of the original instrument. The initial and revised versions of the source language version are then compared, discrepancies are identified, and appropriate revisions are made.

Questionnaire translation, however, may be more art than science, and serious disagreements continue to be raised regarding the efficacy of these traditional procedures to which several generations of students have been introduced. Deutscher (1973) has warned that back-translation "can instill a false sense of security by demonstrating a spurious lexical equivalence," at the expense of interpretive equivalence. Reliance on back-translation may be particularly dangerous for researchers unfamiliar with one or more of the target languages, as these procedures are unlikely to provide critical information regarding the issues underlying translation discrepancies. In this regard, back-translation may be appropriately referred to as a "black box" technique (Harkness, 1996). Other concerns, discussed by Brislin, Lonner and Thorndike (1973) include the fact that, due to their varied backgrounds, translators may not always have an adequate awareness of the methodological requirements of cross-cultural translation, or experience with the subject material they are asked to translate. However, Sperber, DeVellis and Boehlecke (1994) have suggested that highly skilled translators may be successful in developing precise translations of poorly-worded survey questions.

Werner and Campbell (1970) have addressed some of these concerns with their proposal for "decentering" questionnaires. They identify two forms of questionnaire translation: symmetrical and asymmetrical. The basic back-translation process described above is an example of asymmetrical (or unicentered) translation because it emphasizes loyalty to a source language questionnaire that remains unchanged and serves as the standard for the development of target language instruments. Symmetrical (or decentered) translation, in

contrast, may involve multiple iterations of translation and back-translation, with each language version being continually refined to bring them into closer concordance of meaning. This "decentering" approach should be more successful in achieving interpretive equivalence compared to simple back-translation alone.

Another potential approach to addressing the problem of interpretive equivalence in translation is a variation of the back-translation procedure described by Anderson (1967). In essence, he recommends employing groups of bilinguals to work independently to develop a number of alternative versions of both the source and target language instruments. Although costly, this approach would produce a pool of alternate versions of each questionnaire item within which the effects of language, translation, and translator personal idiosyncracies would be random. Use of randomly selected question versions from such a pool and/or the use of different versions with randomly selected subsamples of survey respondents, he suggests, may be one method of producing cross-cultural equivalence.

Sperber, Devellis and Boehlecke (1994) have recently contributed a new step into the translation process in which they quantitatively evaluate source and back-translated questionnaire versions by asking substantive experts (in their example, medical students and faculty) to rank the degree to which the two alternative versions in the source language are comparable. Some practical guidelines for translating psychological tests and instruments have also been recently presented by Van de Vijver and Hambleton (1996).

Another recent innovation in translation research is the development and testing of cognitive thinkaloud protocol translation methodologies by Harkness (1996). The purpose of this approach is to supplement other translation procedures with information regarding how translator's interpret their role, how they approach and perform the task of translation, and the types of information they consider when translating survey questionnaires. Harkness (1996) reports an experiment in which traditional back-

translation procedures were compared with a thinkaloud translation protocol. The procedure was found to contribute a considerable amount of useful information above and beyond that obtained from back-translation alone. This approach should be viewed as an important complement to back-translation, in that it can provide important insights into the reasons for disagreements among translation versions that might otherwise be unavailable to monolingual researchers.

Facet analysis (Canter, 1983) is a related technique that has been recently proposed as a method for improving the development of equivalent survey questions in different languages (Borg, 1996). Consistent with the concept of interpretive equivalence, facet analysis enables one to emphasize shared meaning rather than shared stimulus. This methodology may be useful in identifying the dimensions, or facets, of survey questions. By doing so, questions might be "mapped" into equivalent counterparts in another language without reliance on fallible literal translations. Borg (1996) lists several additional advantages of this technique, including the ability to catalog question types, and to model the conceptual structure of survey questions. He also identifies one important limitation of this approach: the fact that the mapping of survey questions can become very complex, technical and abstract. Translators not expert in a particular substantive area may find such mapping sentences of little help. As mentioned earlier, the lack of substantive knowledge on the part of the translator is a general problem when translating survey instruments. Borgs's paper (this volume) provides an empirical example of how facet analysis might be usefully applied to a questionnaire translation problem.

3.2 Questionnaire Pretesting Phase

Several special techniques for pretesting monocultural survey instruments have also been applied to problems of cross-cultural equivalence. One set of these are *structured probes and/or cognitive interviews*. Schuman's (1966) introduction of the random probing

technique in a cross-cultural setting provides an early example of how follow-up questions can be used to identify respondent difficulties with question interpretation. In his example, responses to these open-ended probes were coded according to the degree to which a subject's response was able to correctly predict their substantive answer to the survey question. More recently, Johnson et al. (1996a; 1997) and Krause and Jay (1994) have employed thinkaloud interviews to examine cross-cultural differences in the cognitive processing of survey questions. Although these techniques are often able to provide important qualitative information that can be used to assess the interpretive equivalence of survey items, there is also the danger that they may interfere with or otherwise influence respondent answers to substantive survey questions. While this risk may be small relative to the potential advantages of cognitive interviewing, it should be recognized, particularly when working with cultural groups that may be unfamiliar with this general methodology.

Another pretesting methodology that has only recently been applied in a cross-cultural setting will be labeled here as *measuring response category intensity*. Unlike most of the other techniques reviewed, which focus on the interpretive equivalence of survey questions, this approach focuses on the interpretation of the response scales used to measure respondent attitudes and opinions. The essential procedure involves asking samples of respondents from multiple cultural groups to assign numeric values to the responses of various classification schemes. Mohler et al. (1996) and Smith (1997) have reported a cross-national experiment recently conducted as part of the ISSP (International Social Survey Programme) that compared the strength of meanings assigned by German and U.S. respondents to the various elements of several commonly employed survey response scales. For example, they evaluated 28 potential response options that reflect various degrees of agreement and disagreement. Smith (1997) concludes that this approach is more advantageous than other potential methods, including simple ranking and magnitude estimation, for measuring the strength of response categories.

Nonetheless, it should be noted that this approach relies on the untested assumption that numeric scales are interpreted in an equivalent manner across cultures.

The *behavior coding* of respondent difficulties in the interpretation of survey items has also been applied to cross-cultural research. Johnson et al. (1996b) employed this technique to examine composite variability in difficulties with interpreting health survey questions across four cultural groups in the U.S.: African Americans, Mexican Americans, Puerto Ricans, and non-Hispanic Whites. More than 300 interviews were tape-recorded and subsequently evaluated to identify respondent behaviors and/or statements that could be reasonably classified as problems relevant to question interpretation (e.g., requests for clarification, inadequate answers). Inkeles and Smith (1974), and Kohn and Slomczynski (1990) have also used behavior coding of pretest data to examine question comprehension problems across cultural groups. Comparative behavioral coding appears to have promise as a method for collecting somewhat more objective evidence of differential interpretation problems across cultures. This procedure, however, rests on the often-questionable assumption of cross-cultural similarities in response styles, such as satisficing (Krosnick, 1991) and courtesy bias (Jones, 1963), which may influence respondent expressions and indications of interpretation difficulty.

A final approach to evaluating cross-cultural equivalence during questionnaire pretesting is to examine respondent answers across *alternative data collection modes*.

This approach is recommended by Flaherty et al. (1988) in order to insure technical equivalence across groups. Although it may often be tempting and convenient to do so, of course, it cannot be assumed that all cultures will react to the same survey methods in an identical manner. Aquilino and LoSciuto (1990), for example, have provided evidence that African American, but not White, respondents may be significantly less likely to report drug use during telephone, compared to in-person, interviews. Unfortunately, although findings such as these have important implications for the collection of cross-

cultural survey data, the mode of data collection is often fixed and questions of cultural differences in mode effects are never considered, let alone addressed.

3.3 Data Collection Phase

Many researchers recommend using ***multiple indicators*** to measure each topic examined in cross-cultural surveys (Braun and Scott, 1996; Mitchell, 1973; Okazaki and Sue, 1995; Przeworski and Teune, 1970; Smith, 1988). Although this recommendation is also relevant to monocultural surveys (Elder, 1976), as it can demonstrably improve measurement quality, it is likely to take on added importance in cross-cultural surveys. This is because post-survey data analyses (see next section) may identify some questions that do not perform in an identical manner (for example, do not cluster in a similar fashion) across cultures. One can therefore avoid "placing-all-of-the-eggs-in-one-basket" by developing multiple survey indicators for each construct to be measured. Smith (1988) suggests using at least three indicators of each construct; items that employ different response scales as well as different questions. These recommendations are very reasonable and should be considered even by those researchers who either: (1) do not have the resources to implement any of the other strategies discussed up to this point; or (2) are "certain" that their own research will be graced with interpretational and procedural equivalence without the need to resort to any of these additional methodologies.

Another approach goes beyond the simple collection of multiple indicators by ***including both etic and emic questions*** in the survey instrument. That is, this procedure asks a set of questions that are thought to have universal relevance across the cultures being surveyed, as well as additional sets believed to be relevant only to some cultures or to have unique meanings across all cultures. This alternative follows the recommendations of both Przeworski and Teune (1970) and Triandis (1972), who have presented

methodologies (to be discussed below) for jointly analyzing both types of questions. It is of further interest because it appears to address both interpretive and procedural equivalence by acknowledging that conceptually identical phenomena may be successfully measured across cultures using different instruments. While this is a powerful approach, it poses significant challenges to researchers. As Frijda and Jahoda (1966) observe, developing survey materials that are appropriate for a given culture makes the often questionnable assumption that the researcher has a detailed and intimate understanding of the culture(s) being studied. Some of the collaborative suggestions discussed earlier may help address this important concern. In addition, as Warwick and Osherson (1973) have observed, because this approach recommends that the emic questions be asked of respondents within each culture, respondents may sometimes be asked to answer survey questions that appear irrelevant or even foolish to them. In order to avoid this latter possibility, investigators may sometimes be inclined to exclude important emic questions from the survey instrument, even at the risk of restricting the relevant question content for one or more cultures.

Another data collection procedure that is commonly employed in hopes of approximating procedural equivalence is *respondent-interviewer matching* on one or more demographic characteristics, although primarily race/ethnicity or gender is taken (Couper, 1991; Schaeffer, 1980), or the use of indigenous interviewers (Bloom and Padilla, 1979). These practices are usually implemented with the expectation that respondents will feel more at ease, and be more forthcoming with their answers, when the perceived social distance between themselves and their interviewer is low. Brislin (1986), for instance, has argued that matching will contribute to the minimization of various types of response bias that may result from the uncertainties of cross-cultural communication. Language problems should also be minimized under these conditions. Hanna and Hanna (1966) have stated that in some societies failure to match respondents with similar interviewers will produce data in which we can have no “confidence.” However, there is not universal agreement on the applicability of matching procedures. Ferketich, Phillips and Verran (1993) have

observed that in communities where the need for privacy may be strong, outside or otherwise dissimilar interviewers may be preferred. Others have argued that a highly trained staff of interviewers who are given random interview assignments is the most effective approach to minimizing response bias (Collins, 1980; Freeman and Butler, 1976).

3.4 Data Analysis Phase

The most basic form of data analysis for assessing one or more forms of cross-cultural equivalence is to employ *item analysis* techniques. At a minimum, researchers should examine frequency distributions for obvious indications of variability across groups, such as differing or high proportions of "don't know" responses, which may indicate lack of interpretive equivalence (Frijda and Jahoda, 1966; Smith, 1988). Likewise, an indicator that lacks variability in one culture but not another is in all likelihood representing an emic concept. Frey (1970) suggests that these types of simple psychometric comparisons may identify the "tip-of-the-iceberg," providing warning of a more serious lack of equivalence hidden below the surface. More elaborate forms of item analysis rely on assessments of cross-group differentials in item functioning using analysis of variance (ANOVA) and other bivariate statistical techniques (Van de Vijver and Leung, 199**??**; Devins et al., 1997).

Other preliminary analysis procedures may examine cross-cultural differences in response styles, such as acquiescence, social desirability, and extreme response style, in an effort to assess the degree to which these variables may be influencing responses from each culture. Another approach is to determine if multiple indicators of each construct correlate with one another in a similar manner across cultural groups. Iyengar (1993) suggests that increased similarity in correlation patterns across groups may be an indicator of interpretive equivalence across groups. Comparisons of scale reliabilities across cultural groups is also used as a preliminary method of investigating procedural

equivalence (Devins et al., 1997). Kuechler (1987) takes a somewhat different approach to item analysis, suggesting that a thorough set of within-group analyses should be completed prior to the conduct of cross-cultural comparisons.

Item response theory methodology is a more sophisticated approach to identifying survey questions that do and do not behave in a similar manner across cultures (Leung and Drasgow, 1986). This technique is commonly used by psychologists to identify test items that do not reflect the underlying latent construct purportedly being measured. Several authors have provided useful examples of the application of item response models to assessments of the translation equivalence (Candell and Hulin, 1987; Ellis et al., 1989; Hulin, 1987) and cross-cultural relevance (Hui et al., 1983) of survey scale items. It does so by comparing cross-group item characteristic curves, which represent the conditional probabilities of responding in a given manner to individual questions for various levels of a latent variable represented by a measurement scale. Similar item characteristic curves across cultural groups are interpreted as evidence of similar behavior, and hence equivalence. Several limitations of this approach have been noted, including the very strong assumption that the underlying latent trait represented by the survey items is unidimensional, an assumption that may seldom be realistic (Hulin et al., 1982). In addition, these models require fairly large numbers of items in order to function properly, also often an unrealistic assumption for many survey data sets, and the requirement that all observed variables be measured on a dichotomous scale (Drasgow and Kanfer, 1985). This methodology, however, does have the ability to incorporate both etic and emic questions into cross-cultural measures (de Vera, 1985; Hulin, 1987), and should thus be considered an option whenever practical.

Another analytic method that is used to evaluate the equivalence of translated instruments is based on *generalizability theory* (Van de Vijver and Poortinga, 1982). Using an analysis of variance framework, this procedure can partial out variability in survey responses due to the effects of language, individuals, other variables, and all interactions

(Hulin, 1987). Katerberg et al. (1977) provide an application of generalizability theory to an evaluation of the equivalence of English and Spanish versions of two job attitude measures. A unique advantage of this method is its potential to view cross-cultural equivalence as a relative, rather than an absolute, concept (Van de Vijver and Poortinga, 1982). A potential limitation of generalizability theory models are their reliance on the responses of bilingual respondents, who are asked to complete the survey instrument in each language. This necessary reliance on bilinguals is an obvious concern because they may not be representative of the monolingual populations that many researchers are more interested in generalizing to. This technique also assumes that the bilingual respondents will answer in a similar manner in either language, a questionable assumption at best. Marin et al. (1983), for instance, found that Spanish-English bilinguals use more complex cognitive structures when completing the questionnaire in their native language.

Several techniques have also been used to compare the structural relationships among sets of survey items across two or more cultural groups. One of these is *confirmatory factor analysis*. This procedure was introduced by Joreskog (1971), who described it as a theory-driven tool that could be used to compare simultaneously the factor structure of a set of survey questions across multiple population groups and make assessments of their equivalence through comparisons of large sample chi-square statistics. There are numerous excellent examples of the application of confirmatory factor analysis to equivalence problems in cross-cultural research (Devins et al., 1997; Drasgow and Kanfer, 1985; Kohn et al., 1997; Miller et al., 1981; Watkins, 1989). Singh (1995) describes several increasingly precise levels of procedural equivalence that can be obtained using this technique. Unlike the item response models discussed earlier, confirmatory factor analysis is useful in examining the relatively small numbers of items that might be available to represent a given construct in many survey questionnaires. Another advantage of confirmatory factor analysis is its ability to take full advantage of the information available in ordinal and interval rating scales, unlike item response theory models which require dichotomous data (Drasgow and Kanfer, 1985). Kuechler (1987),

however, correctly observes that confirmatory factor analysis requires a large number of assumptions that information collected using survey methodologies are often unable to meet.

Multidimensional scaling has additionally been employed to compare the structure of survey measures cross-culturally (Allerbeck, 1977). This technique examines the relative proximities among sets of survey measures to identify their underlying structure. In practice, multidimensional scaling often produces findings similar to factor analyses, although the latter technique permits more rigorous comparisons of alternative models (Van de Vijver and Leung, 199**??**). Schwartz and Sagiv (1995) and Braun and Scott (1996) have utilized multidimensional scaling to conduct cross-cultural comparisons of the dimensionality of survey instruments. Hayashi et al. (1992) report cross-cultural comparisons using a similar technique which they refer to as *minimum dimension analysis*.

Other analytic approaches have also been used to establish procedural equivalence between samples when investigating cultural effects in survey research. One basic approach has been to examine the effects of culture after first *applying statistical controls* for other sources of variation that might be confounded with culture, such as socioeconomic status (cf., Johnson et al., 1997). Another strategy has been suggested by Leung (1989) and Van de Vijver and Leung (forthcoming), who have observed that the concept of culture is far too broad and complex to serve as an acceptable explanatory variable. They suggest that the analyses of survey data collected across cultures may be improved if an approach is adopted that replaces the commonly used global indicators of culture, such as race, ethnicity, and country of origin, with more specific measures that represent the qualities or features of various cultures that are believed to account for the cross-group differences of interest. This strategy is known as the "unpackaging" of culture (Whiting, 1976). A related procedure has been demonstrated by Johnson et al. (1996a), who provide empirical examples of how variability in survey question

interpretation may be able to account for cultural differences in self-reported physical and mental health. Poortinga (1989) has referred to this approach as "interpreting equivalence."

One final approach to establishing cross-cultural equivalence when analyzing survey data is what Przeworski and Teune (1970) have referred to as the *identity-equivalence method*. Briefly, this method would include survey instrument items that are thought to be etic across each of the cultures of interest, as well as questions believed to be emic to one or some of the cultures being examined (see above). A subsequent set of statistical analyses using correlation matrices, factor analysis or some other technique, would be used to verify empirically which measures were representing the same construct cross-culturally. Survey questions not identified as etic may nonetheless be valid emic indicators of the construct being examined if they correlate with the etic items within a given culture. The measure of an etic construct may thus be developed using a common set of emic indicators and group-specific sets of emic items. A important feature of this approach is its attempt to reconcile interpretive and procedural equivalence. It should be noted that this procedure is similar to the concept of "etic + emic" analysis outlined by Harry Triandis and colleagues (Davidson et al., 1976; Triandis, 1972; Triandis and Marin, 1983). Kohn and Slomczynski (1990) provide an excellent example of the application of the identity-equivalence method as part of their comparative analyses of the relationship between social structure and personality in Poland and the U.S. Examples of other studies that have employed this technique include Funkhouser (1993), Miller et al. (1985), Przeworski and Teune (1966), and Verba et al. (1978). A disadvantage of this approach is its seeming inability to be used in conjunction with pooled analyses of cross-cultural data sets.

4. Discussion

In addition to the traditional reliability and validity requirements for monocultural survey instruments, researchers conducting cross-cultural survey research have the added concern of equivalence. Indeed, cross-cultural research demands a commitment to the establishment of equivalence that is at least equal to the attention routinely reserved for the problems of reliability and validity. As this review suggests, cross-cultural equivalence has been conceptualized in a multitude of ways, and social scientists have in turn devised a variety of methods for use in hopes of achieving it. Although equivalence has multiple dimensions, there seems to be a natural distinction between interpretive and procedural equivalence. While interpretive equivalence is primarily concerned with the subjective cross-cultural comparability of meaning, procedural equivalence, broadly speaking, refers to the objective development of comparable survey measures across cultural groups. Depending on the research questions of interest, the various dimensions of equivalence represented by these two general labels may take on different levels of importance.

It should also be noted that not all forms of equivalence are necessarily created equal. Whereas most would agree that interpretational equivalence is an absolute requirement, certain forms of procedural equivalence may not always be necessary, or even desirable. Specific forms of procedural equivalence that emphasize pure replication of survey questions across cultures may, for example, be inappropriate in many situations where differing norms or frames of reference may require unique survey measures of the same construct. Nonetheless, many otherwise conscientious researchers prefer working with identically-worded survey questions in cross-cultural studies, even when evidence of poor interpretational equivalence is readily available, because such procedural equivalence facilitates data analysis. Indeed, the challenges that an emphasis on interpretational equivalence can pose for data analysis is likely the main reason why so many cross-cultural studies prefer to emphasize forms of procedural equivalence instead. The general

underdevelopment of cross-cultural survey research methodology mentioned earlier in this paper can probably be attributed to this expediency more than anything else.

Ironically, despite this state of affairs, numerous methods for establishing or assessing one or more forms of cross-cultural equivalence are currently available. The best advice to researchers is probably to employ as many of these techniques as possible and within reason, given that various methodologies may be more appropriate to one specific form of equivalence or another. Several other researchers, including Hui and Triandis (1985), and Van de Vijver and Poortinga (1997), have made similar recommendations. Certainly, efforts to establish cross-cultural equivalence should be made during each phase of survey implementation. Various forms of interpretive equivalence, for example, can be and are more easily assessed during question development and questionnaire pretesting phases, while issues of procedural equivalence tend to predominate during the data collection and analysis stages. One gross indicator of the success researchers have had in establishing cross-cultural equivalence may simply be the number of alternative methods they employed throughout the course of their study to achieve this goal.

Finally, efforts to improve the available tools for developing cross-cultural equivalence should be recognized now as one of the more pressing needs of the survey research community. As the cultural composition of many countries continues to diversify, an ever increasing proportion of all researchers will need to confront issues of equivalence in the conduct of their work. The international research community would be the beneficiary if all graduate programs and survey research centers emphasized the importance of cross-cultural equivalence and encouraged ongoing theoretical and methodological assessments of this fundamental problem.

References

Adams-Esquivel, H. (1991) Conceptual adaptation vs. Back-translation of multilingual instruments: How to increase the accuracy and actionability of multlingual surveys. Paper resented at the annual meeting of the American Association for Public Opinion Research, Pheonix, AZ.

Allerbeck, K.R. (1977) Analysis and inference in cross-national survey research. Pp. 373-402 in A. Szalai and R. Petrella (Eds.) Cross-National Comparative Survey Research: Theory and Practice. Oxford: Pergamon.

Alwin, D.F., Braun, M., Harkness, J. and Scott, J. (1994) Measurement in multi-national surveys. Pp. 26-39 in I. Borg and P. Mohler (Eds.) Trends and Perspectives in Empirical Social Research. Berlin: Walkter de Gruyter.

Anderson, R.B. (1967) On the comparability of meaningful stimuli in cross-cultural research. Sociometry 30: 124-136.

Aquilino, W. and LoSciuto, L. (1990) Effects of interview mode on the validity of drug use surveys." Public Opinion Quarterly 54: 362-395.

Barry, H. (1980) Description and uses of the Human Relations Area Files. Pp. 445-478 in H.C. Triandis and J.W. Berry (Eds.) Handbook of Cross-Cultural Psychology: Methodology, Volume 2. Boston: Allyn and Bacon.

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

Berry, J.W., Poortinga, Y.H., Segall, M.H. and Dasen, P.R. (1992) Cross-Cultural Psychology: Research and Applications. New York: Cambridge University Press.

Bernard, H.R. (1988) Research Methods in Cultural Anthropology. Newbury Park, CA: Sage.

Bloom, D. And Padilla, A.M. (1979) A peer interviewer model in conducting surveys among Mexican-American youth. Journal of Community Psychology 7: 129-136.

Blumer, M. and Warwick, D.P. (1993) Social Research in Developing Countries: Surveys and Censuses in the Third World. London: John Wiley & Sons.

Borg, I. (1996) Using facet theory to control item content in cross-cultural surveys. Paper presented at the International Sociological Association Conference on Social Science Methodology, University of Essex, Colchester, England.

Borg, I. And Shye, S. (1995) Facet Theory: Form and Content. Thousand Oaks, CA: Sage.

Braun, M. and Scott, J. (1996) Data-based procedures for the detection of problems in functional equivalence - or: Is "having a job" the same as "working"? Paper presented at the International Sociological Association Conference on Social Science Methodology, University of Essex, Colchester, England.

Brislin, R.W. (1970) Back-translation for cross-cultural research. Journal of Cross-Cultural Psychology 1: 195-216.

Brislin, R.W. (1976) Translation: Applications and Research. New York: John Wiley.

Brislin, R.W. (1986) The wording and translation of research instruments. Pp. 137-164 in Lonner, W.J. and Berry, J.W. (Eds.) Field Methods in Cross-Cultural Research. Beverly Hills, CA: Sage.

Brislin, R.W., Lonner, W.J. and Thorndike, R.M. (1973) Cross-Cultural Research Methods. New York: John Wiley & Sons.

Candell, G.L. and 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.

Collins, M. (1980) Interviewer variability: A review of the problem. Journal of he Market Research Society 22: 77-95.

Couper, M.P. (1991) Modeling survey participation at the interviewer level. Proceedings of the Section on Survey Research Methods, American Statistical Association, 98-107.

Czudnowski, M.M. (1976) Comparing Political Behavior. Beverly Hills, CA: Sage.

Davidson, A.R., Jaccard, J.J., Triandis, H.C., Morales, M.L. and Diaz-Guerrero, R. (1976) Cross-cultural model testing: Toward a solution of the etic-emic dilemma. International Journal of Psychology 11: 1-13.

de Vera, M.V. (1985) Establishing cultural relevance and measurement equivalence using emic and etic items. Unpublished dissertation. Urbana, IL: University of Illinois.

Deutscher, I. (1973) Asking questions cross-culturally: Some problems of linguistic comparability. Pp. 163-186 in D.P. Warwick and S. Osherson (Eds.) Comparative Research Methods. Englewood Cliffs, NJ: Prentice-Hall.

Devins, G.M., Beiser, M., Dion, R., Pelletier, L.G. and Edwards, R.G. (1997) Cross-cultural measurements of psychological well-being: The psychometric equivalence of Cantonese, Vietnamese, and Laotian translations of the affect balance scale. American Journal of Public Health 87: 794-799.

Drasgow, F. and Kanfer, R. (1985) Equivalence of psychological measurement in heterogeneous populations. Journal of Applied Psychology 70: 662-680.

Dressler, W.M., Viteri, F.E., Chavez, A., Grell, G.A.C. and Dos Santos, J.E. (1991) Comparative research in social epidemiology: Measurement issues. Ethnicity and Disease 1: 379-393.

Dunnigan T., McNall, M. and Mortimer, J.T. (1993) The problem of metaphorical nonequivalence in cross-cultural survey research. Journal of Cross-Cultural Psychology 24: 344-365.

Eckensberger, L.H. (1973) Methodological issues of cross-cultural research in developmental psychology. Pp. 43-64 in J.R. Nesselroade and H.W. Reese (Eds.) Life-Span Developmental Psychology: Methodological Issues. New York: Academic Press.

Elder, J.W. (1973) Problems of cross-cultural methodology: Instrumentation and interviewing in India. Pp. 119-144 in M. Armer and A.D. Grimshaw (Eds.) Comparative Social Research: Methodological Problems and Strategies. New York: John Wiley & Sons.

Elder, J.W. (1976) Comparative cross-national methodology. Pp. 209-230 in A. Inkeles, J. Coleman and N. Smelser (Eds.), Annual Review of Sociology, Volume 2. Palo Alto, CA: Annual Reviews, Inc.

Ellis, B.B., Minsel, B. and Becker, P. (1989) Evaluation of attitude survey translations: An investigation using item response theory. International Journal of Psychology 24: 665-684.

Eyton, J. And Neuwirth, G. (1984) Cross-cultural validity: Ethnocentrism in health studies with special reference to the Vietnamese. Social Science and Medicine 5: 447-453.

Ferketich, S., Phillips, L. and Verran, J. (1993) Development and administration of a survey instrument for cross-cultural research. Research in Nursing & Health 16: 227-230.

Flaherty, J.A., Gaviria, M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J.A. and Birz, S. (1988) Developing instruments for cross-cultural psychiatric research. Journal of Nervous and Mental Disease 176: 257-263.

Freeman, J. and Butler, E.W. (1976) Some sources of interviewer variance in surveys. Public Opinion Quarterly 40: 79-92.

Frey, F.W. (1970) Cross-cultural survey research in political science. Pp. 173-294 in Holt, R.T. and Turner, J.E. (Eds.) The Methodology of Comparative Research. New York: Free Press.

Frijda, N. and Jahoda, G. (1966) On the scope and methods of cross-cultural research. International Journal of Psychology 1: 109-127.

Funkhouser, G.R. (1993) A self-anchoring instrument and analytical procedure for reducing cultural bias in cross-cultural research. The Journal of Social Psychology 133: 661-673.

Green, R.T. and White, P.D. (1976) Methodological considerations in cross-national consumer research. Journal of International Business Studies 7: 81-87.

Hanna, W.J. and Hanna, J.L. (1966) The problem of ethnicity and factionalism in African survey research. Public Opinion Quarterly 30: 290-294.

Harari, O. and Beaty, D. (1990) On the folly of relying solely on a questionnaire methodology in cross-cultural research. Journal of Managerial Issues 11: 267-281.

Harkness, J. (1996) Cognitive approaches to survey translation. Paper presented at the International Sociological Association Conference on Social Science Methodology, University of Essex, Colchester, England.

Hayashi, C., Suzuki, T. and Sasaki, M. (1992) Data Analysis for Comparative Social Research: International Perspectives. Amsterdam: North-Holland.

Henderson, D.J., Sampselle, C., Mayes, F. and Oakley, D. (1992) Toward culturally sensitive research in a multicultural society. Health Care for Women International 13: 339-350.

Hines, A.M. (1993) Linking qualitative and quantitative methods in cross-cultural survey research: Techniques from cognitive science. American Journal of Community Psychology 21: 729-746.

Hui, C.H., Drasgow, F. and Chang, B.H. (1983) An analysis of the modernity scale: An item response theory approach. Journal of Cross-Cultural Psychology 14: 259-278.

Hui, C.H. and Triandis, H.C. (1983) Multistrategy approach to cross-cultural research: The case of locus of control. Journal of Cross-Cultural Psychology 14: 65-83.

Hui, C.H. and Triandis, H.C. (1985) Measurement in cross-cultural psychology: A review and comparison of strategies. Journal of Cross-Cultural Psychology 156: 131-152.

Hulin, C.L. (1987) A psychometric theory of evaluations of item and scale translations: Fidelity across languages. Journal of Cross-Cultural Psychology 18: 115-142.

Hulin, C.L., Drasgow, R. and Comocar, J. (1982) Applications of item response theory to analysis of attitude scale translations. Journal of Applied Psychology 67: 818-825.

Inkeles, A. and Smith, D.H. (1974) Becoming Modern: Individual Change in Six Developing Countries. Cambridge, MA: Harvard University Press.

Iyengar, S. (1993) Assessing linguistic equivalence in multilingual surveys. Pp. 173-182 in M. Blumer and D.P. Warwick (Eds.) Social Research in Developing Countries: Surveys and Censuses in the Third World. London: John Wiley & Sons.

Johnson, T.P., Jobe, J., O'Rourke, D., Sudman, S., Warnecke, R., Chavez, N., Chapa-Resendez, G. and Golden, P. (forthcoming) Dimensions of identification among multiracial and multiethnic respondents in survey interviews. Evaluation Review.

Johnson, T., O'Rourke, D., Chavez, N., Sudman, S., Warnecke, R., Lacey, L. and Horm, J. (1996a) Cultural variations in the interpretation of health survey questions. Pp. 57-62 in Warnecke, R. (Ed.) Health Survey Research Methods Conference Proceedings. DHHS Publication No. (PHS) 96-1013. Hyattsville, MD: National Center for Health Statistics.

Johnson, T.P., O'Rourke, D., Sudman, S., Warnecke, R. and Chavez, N. (1996b) Assessing question comprehension across cultures: Evidence from the United States. Paper presented at the International Sociological Association Conference on Social Science Methodology, University of Essex, Colchester, England.

Johnson, T., O'Rourke, D., Chavez, N., Sudman, S., Warnecke, R., Lacey, L. and Horm, J. (1997) Social cognition and responses to survey questions among culturally diverse populations. Pp. 87-113 in Lyberg, L., Biemer, P., Collins, M., de Leeuw, E., Dippo, C., Schwarz, N. and Trewim, D. (Eds.) Survey Measurement and Process Quality. New York: John Wiley & Sons.

Jones, E.L. (1963) The courtesy bias in South-East Asian surveys. International Social Science Journal 15: 70-76.

Jones, E.G. and Kay, M. (1992) Instrumentation in cross-cultural research. Nursing Research 41: 186-188.

Joreskog, K.G. (1971) Simultaneous factor analysis in several populations. Psychometrika 36: 409-426.

Katerberg, R., Smith, F.J. and Hoy, S. (1977) Language, time, and person effects on attitude scale translations. Journal of Applied Psychology 62: 385-391.

Kleinman, A. (1987) Anthropology and psychiatry: The role of culture in cross-cultural research on illness. British Journal of Psychiatry 151: 447-454.

Kohn, M.L. and Slomczynski, K.M. (1990) Social Structure and Self-Direction: A Comparative Analysis of the United States and Poland. Cambridge, MA: Basil Blackwell.

Kohn, M.L., Slomczynski, K.M., Janicka, K., Khmelko, V., Mach, B.W., Paniotto, V., Zaborowski, W., Gutierrez, R. and Heyman, C. (1997) Social structure and

personality under conditions of radical social change: A comparative analysis of Poland and Ukraine. American Sociological Review 62: 614-638.

Krause, N.M. and Jay, G.M. (1994) What do global self-rated health items measure? Medical Care 32: 930-942.

Krosnick, J.A. (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied Cognitive Psychology 5: 213-236,

Kuechler, M. (1987) The utility of surveys for cross-national research. Social Science Research 16: 229-244.

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

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

Leung, K. And Drasgow, F. (1986) 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.

Marin, G. and Marin, B. (1991) Research with Hispanic Populations. Newbury Park, CA: Sage.

Marin, G., Triandis, H.C., Betancourt, H. And Kashima, Y. (1983) Ethnic affirmation versus social desirability: Explaining discrepancies in bilinguals' responses to a questionnaire. Journal of Cross-Cultural Psychology 14: 173-186.

Marsh, R.M. (1967) Comparative Sociology: A Codification of Cross-Societal Analysis. New York: Harcourt, Brace & World.

McKay, R.B., Breslow, M.J., Sangster, R.L., Gabbard, S.M., Reynolds, R.W., Nakamoto, J.M. and Tarnai, J. (1996) Translating survey questionnaires: Lessons learned. New Directions for Evaluation 70: 93-104.

Miller, J., Slomczynski, K.M. and Kohn, M.L. (1985) Continuity of learning generalization: The effect of job on men's intellective process in the United States and Poland. AJS 91: 593-615.

Miller, J., Slomczynski, K.M. and Schoenberg, R.J. (1981) Assessing comparability of measurement in cross-national research: Authoritarian-conservatism in different sociocultural settings. Social Psychology Quarterly 44: 178-191.

Mitchell, R.E. (1973) Survey materials collected in the developing countries: Sampling, measurement, and interviewing obstacles to intra- and inter-national comparisons. Pp. 204-226 in D.P. Warwick and S. Osherson (Eds.) Comparative Research Methods. Englewood Cliffs, NJ: Prentice-Hall.

Mohler, P., Harkness, J., Smith, T.W. and Davis, J.A. (1996) Calibrating response scales across two languages and cultures. Paper presented at the International Sociological Association Conference on Social Science Methodology, University of Essex, Colchester, England.

Mullen, M.R. (1995) Diagnosing measurement equivalence in cross-national research. Journal of International Business Studies 36: 573-596.

Narula, U. (1990) Practical constraints in social field research in India. Pp. 123-149 in U. Narula and W.B. Pearce (Eds.) Culturs, Politics, and Research Programs: An International Assessment of Practical Problems in Field Research. Hillsdale, NJ: Lawrence Erlbaum Associates.

NieBen, M. (1982) Qualitative aspects in cross-national comparative research and the problem of functional equivalence. Pp. 83-104 in M. Niessen and J. Peschar (Eds.) International Comparative Research. Oxford: Pergamon Press.

Okazaki, S. and Sue, S. (1995) Methodological issues in assessment research with ethnic minorities. Psychological Assessment 7: 367-375.

Pareek, U. and Rao, T.V. (1980) Cross-Cultural surveys and interviewing. Pp. 127-179 in H.C. Triandis and J.W. Berry (Eds.) Handbook of Cross-Cultural Psychology, Volume 2. Boston: Allyn and Bacon.

Peschar, J. (1982) Quantitative aspects in cross-national comparative research: Problems and issues. Pp. 57-81 in M. Niessen and J. Peschar (Eds.) International Comparative Research. Oxford: Pergamon Press.

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

Prince, R. and Mombour, W. (1967) A technique for improving linguistic equivalence in cross-cultural surveys. Journal of Social Psychology 13: 229-237.

Przeworski, A. and Teune, H. (1966) Equivalence in cross-national research. Public Opinion Quarterly 30: 33-43.

Przeworski, A. and Teune, H. (1970) The Logic of Comparative Social Inquiry. New York: John Wiley & Sons.

Schachter, S. (1954) Interpretive and methodological problems of replicated research. Journal of Social Issues 10: 52-60.

Schaeffer, N.C. (1980) Evaluating race-of-interviewer effects in a national survey. Sociological Methods & Research 8: 400-419.

Scheuch, E.K. (1993) The cross-cultural use of sample surveys: Problems of comparability. Historical Social Research 18: 104-138.

Schuman, H. (1966) The random probe: A technique for evaluating the validity of closed questions. American Sociological Review 31: 218-222.

Schwartz, S.H. and Sagiv, L. (1995) Identifying culture-specifics in the content and structure of values. Journal of Cross-Cultural Psychology 26: 92-116.

Sears, R.R. (1961) Transcultural variables and conceptual equivalence. Pp. 445-455 in B. Kaplan (Ed.) Studying Personality Cross-Culturally. Evanston, IL: Row, Peterson and Co.

Sechrest, L., Fay, T.L. and Hafeez Zaidi, S.M. (1972) Problems of translation in cross-cultural research. Journal of Cross-Cultural Psychology 3: 41-56.

Sekaran, U. (1983) Methodological and theoretical issues and advancements in cross-cultural research. Journal of International Business Studies 14: 61-73.

Singh, J. (1995) Measurement issues in cross-national research. Journal of International Business Studies 26: 597-619.

Smith, T.W. (1988) The ups and downs of cross-national survey research. GSS Cross-National Report No. 8. National Opinion Research Center, University of Chicago.

Smith, T.W. (1997) Improving cross-national survey research by measuring the intensity of response categories. GSS Cross-National Report No. 17. National Opinion Research Center, University of Chicago.

Sperber, A.D., DeVellis, R.F. and Boehlecke, B. (1994) Cross-cultural translation: Methodology and Validation. Journal of Cross-Cultural Psychology 25: 501-524.

Straus, M.A. (1969) Phenomenal identity and conceptual equivalence of measurement in cross-national comparative research. Journal of Marriage and the Family 81: 233-239.

Teune, H. (1977) Analysis and interpretation in cross-national survey research. Pp. 95-128 in A. Szalai and R. Petrella (Eds.) Cross-National Comparative Survey Research: Theory and Practice. Oxford: Pergamon.

Teune, H. (1990) Comparing countries: Lessons learned. Pp. 38-62 in Comparative Methodology: Theory and Practice in International Social Research. London: Sage.

Triandis, H.C. (1972) The Analysis of Subjective Culture. New York: John Wiley & Sons.

Triandis, H.C. (1977) Interpersonal Behavior. Monterey, CA: Brooks/Cole.

Triandis, H.C. and Marin, G. (1983) Etic plus emic versus pseudoetic: A test of a basic assumption of contemporary cross-cultural psychology. Journal of Cross-Cultural Psychology 14: 489-500.

Van de Vijver, F. and Hambelton, R.K. (1996) Translating tests: Some practical guidelines. European Psychologist 1: 89-99.

Van de Vijver, F. and Leung, K. (1996) Methods and data analysis of comparative research. Pp. 257-300 in J.W. Berry, Y.H. Poortinga and J. Pandey (Eds.) Handbook of Cross-Cultural Psychology, Second Edition, Volume 1. Chicago: Allyn & Bacon.

Van de Vijver, F. and Leung, K. (Forthcoming) Methods and Data Analysis for Cross-Cultural Research. Thousand Oaks, CA: Sage.

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

Van de Vijver, F.R. and Poortinga, Y.H. (1997) Towards an integrated analysis of bias in cross-cultural assessment. European Journal of Psychological Assessment, forthcoming.

Verba, S. (1969) The uses of survey research in the study of comparative politics: Issues and strategies. Pp. 56-106 in S. Rokkan, S. Verba, J. Viet, and E. Almasy (Eds.) Comparative Survey Analysis. Paris: Mouton.

Verba, S., Nie, N.H. and Kim, J.O. (1978) Participation and Political Equality: A Seven-Nation Comparison. London: Cambridge University Press.

Warwick, D.P. and Osherson, S. (1973) Comparative Research Methods. Englewood Cliffs, NJ: Prentice-Hall.

Watkins, D. (1989) The role of confirmatory factor analysis in cross-cultural research. International Journal of Psychology 24: 685--701.

Webster's Seventh New Collegiate Dictionary (1965) Springfield, MA: G. & C. Merriam Co.

Werner, L. And Campbell, D.T. (1970) Translating, working through interpreters and the problem of decentering. Pp 398-420 in R. Naroll and R. Cohen (Eds.) American Handbook of Methods in Cultural Anthropology. Garden City, NY: Natural History Press.

Whiting, B.B. (1976) The problem of the packaged variable. Pp. 303-309 in K. Riegel and J. Meacham (Eds.) The Developing Individual in a Changing World, Volume 1. The Hague: Mouton.

Word, C.O. (1992) Cross-cultural methods for survey research in Black urban areas. Pp. 28-42 in Burlew, A.K.H., Banks, W.C., McAdoo, H.P. and Azibo, D.A. (Eds.) African American Psychology: Theory, Research, and Practice. Newbury Park, CA: Sage.