DOES INTERCULTURAL SENSITIVITY CROSS CULTURES? AN INVESTIGATION OF VALIDITY ISSUES INVOLVED IN PORTING INSTRUMENTS ACROSS LANGUAGES AND CULTURES

by

JOE GREENHOLTZ

B.A. University of Winnipeg, 1978
M.Ed. Temple University, 1986

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF EDUCATION

in

THE FACULTY OF GRADUATE STUDIES

(Educational Leadership and Policy)

THE UNIVERSITY OF BRITISH COLUMBIA

July 2005

© Joe Greenholtz, 2005
This study examined the validity of data obtained using a Japanese translation of version one of the Intercultural Development Inventory (IDI) (Bennett & Hammer, 1998) and, by extension, some assumptions surrounding the cross-cultural transferability of the Developmental Model of Intercultural Sensitivity (DMIS) upon which it was based.

First, the research was situated within the field of international exchanges in education. It explored what internationalisation means in an educational context, and how it might be conceptualised and measured as an outcome.

The case was made that an increase in intercultural sensitivity is a desirable, enduring, and achievable legacy for students to take home with them.

The difference between intercultural sensitivity and intercultural competence or the psychological traits posited to underlie intercultural adaptability were discussed and a rationale given for using the Developmental Model of Intercultural Sensitivity (DMIS) as the theoretical framework that best represents the construct.

The protocol for translating the IDI was reported and issues related to translating and adapting instruments were raised. The translation protocol used for this study was a
five-step procedure, much more rigorous in nature than the widely-used translation/back-translation protocol. The added step of having translators use the draft translation to perform a task outside the translation context was a contribution to the art of translating and adapting instruments.

The validation procedure followed Messick's (1998) conception of validity as a multi-layered process incorporating both statistical and judgmental evidence. Examinations of content, concurrent, and consequential validity took place within Messick's (1998) framework positing these aspects of validity to be contributors to an overarching judgment of construct validity. Results of the validation process raised questions about whether the concepts that comprise the IDI, and by extension the DMIS, are transferable across languages and cultures.

These questions regarding transferability of the IDI and the DMIS were discussed in detail and implications for practitioners using existing instruments in a cross-cultural context, and in the context of validity issues, were elucidated.

Finally, the limitations of the study and directions for future research were outlined.
## TABLE OF CONTENTS

ABSTRACT ......................................................................................... ii

TABLE OF CONTENTS ........................................................................ iv

LIST OF TABLES ................................................................................ vi

LIST OF FIGURES ............................................................................... vii

ACKNOWLEDGEMENTS ...................................................................... viii

CHAPTER ONE OVERVIEW AND INTRODUCTION ................................. 1

1.1 OVERVIEW .................................................................................. 1

1.2 INTRODUCTION ........................................................................... 4

1.3 DEFINING THE FIELD ................................................................. 9

1.4 A FRAMEWORK FOR INTERCULTURAL SENSITIVITY ................. 14

1.5 MEASURING INTERCULTURAL SENSITIVITY ............................. 21

1.6 LINGUISTIC DETERMINISM ...................................................... 24

CHAPTER TWO METHODOLOGY ....................................................... 27

2.1 INTRODUCTION ........................................................................... 27

2.2 CONTENT VALIDITY .................................................................... 29

2.3 PROCEDURES FOR TRANSLATING THE IDI ............................. 30

2.4 ASSESSING THE ADEQUACY OF CONSTRUCT REPRESENTATION .... 37

2.4.1 Participants .............................................................................. 37

2.4.2 Principal Components Analysis ............................................ 38
LIST OF TABLES

Table 2.1 Translation Protocol................................. 36
Table 3.1 Component Correlation Matrix......................... 40
Table 3.2 KMO Measure of Sampling Adequacy and Bartlett's Test of Sphericity.............................. 41
Table 3.3 Rotated Factor Matrix................................ 43
Table 3.4 Cronbach's Coefficient Alpha.......................... 52
Table 4.1 Correlations between IDI and Worldmindedness and Intercultural Anxiety Scales....................... 69
Table 4.2 Analysis for Gender Effect............................. 72
LIST OF FIGURES

Figure 1.1 Continuum of Intercultural Sensitivity............17

Figure 3.1 Scree Plot...........................................42
ACKNOWLEDGEMENTS

I would like to thank Drs Milton Bennett and Mitchell Hammer for granting me permission to translate the Intercultural Development Inventory into Japanese and to use it in my research.

I would also like to thank my research assistants, Vickie Yau, who kept all of my data in good order and was always there when I needed to be reminded how to find them, and Michi Ohashi, whose insights into language and cultural concepts helped shape my thinking about the issues involved in porting an instrument from one culture into another.

I am extremely grateful to my translation team, Makoto Morise, Yumiko Morise, Naoko Robb, and Shizu Yamamoto who contributed their expertise and their time, and to the raters, Masaki Kobayashi and Mitsunori Takakuwa whose insights helped me to refine and shape the final version.

I would like to thank the members of my committee, Dan Pratt, the chair, Bruno Zumbo, and Pierre Walter whose incisive comments kept me focussed and who encouraged me to believe in the value of the work I was doing.

I am especially grateful to David Coulter, Tom Sork and the other champions of the Ed D cohort programme for creating this opportunity to re-acquaint, at least this semi-jaded
practitioner, with the world of ideas and to see practice in a
new and invigorating light.

Finally, I would like to thank my wife Hiroko and my
family for putting up with my distracted air, for being
enthusiastic and supportive about the challenge I was
undertaking, and for not rolling their eyes too much during my
repeated attempts to answer the seemingly innocuous question,
'So, what's your research about?'
CHAPTER ONE
OVERVIEW AND INTRODUCTION

1.1 Overview

The purpose of this study was to examine the validity of data obtained using a Japanese translation of version one of the Intercultural Development Inventory (IDI) (Bennett & Hammer, 1998) and, by extension, some assumptions surrounding the cross-cultural transferability of the Developmental Model of Intercultural Sensitivity (DMIS) upon which it is based.

This dissertation situated this research within the field of international exchanges in education. It explored, briefly, what internationalisation means in an educational context, and how it might be conceptualised and measured as an outcome.

I examined the goals that students and institutions have for a year abroad or other international exchange experiences and made the case that an increase in intercultural sensitivity is a desirable, enduring, and achievable legacy for students to take home with them.

The construct of intercultural sensitivity is related to, but fundamentally different from, intercultural competence or the psychological traits posited to underlie intercultural adaptability. There was a discussion of those differences in order to demonstrate that the Developmental Model of
Intercultural Sensitivity (DMIS), the theoretical framework that was chosen, best represents the construct.

The Intercultural Development Inventory (IDI) is the psychometric instrument that operationalises and measures the DMIS. Because the IDI was available only in English when I embarked on this study, it was first necessary to translate it into Japanese so that I could use it with my intended student population. However, the translation process raised many questions about whether the concepts that comprised the instrument, and by extension the DMIS, were transferable across language and culture. In their writings and in IDI qualifying workshops, the instrument’s authors, Bennett and Hammer, always characterise the instrument as reliable and valid. In fact, there was some surprise when I informed them I wanted to do a validation study of the instrument in translation, because ‘it’s already been validated’ (personal correspondence, 2003). The fact that validity is not a property of an instrument, but rather of the inferences one can draw from the data it yields in a particular context, necessitated the empirical investigation that comprised this study.

I reported on the protocol used to translate the IDI, and on the validation process. After outlining the results, I also discussed the implications of this study for validity.
assessment in a cross-cultural context, in terms of translation protocols, in terms of the cross-cultural transferability of the DMIS and the IDI, and as these issues affect practitioners looking for research tools to use in cross-cultural investigations.
1.2 Introduction

This study was to comprise the first stage of a larger research project intended to examine one facet of my practice as Executive Director of a bilateral academic programme, now in its fourteenth year, between a major Canadian research university and a leading private Japanese university. The aim of the larger research agenda was to assess how well the exchange programme was fulfilling its mandate, as set out in its Mission Statement, to 'develop intercultural understanding among the students'. However, the results obtained led me to temporarily curtail the rest of the research because of serious issues they raised regarding the cross-cultural transferability of the theoretical framework and the validity of the research instrument I had used, in translation.

The Mission Statement calls on the academic exchange programme (AEP), "to contribute to the internationalisation of the partner universities, and to develop intercultural understanding among the students through an academic, linguistic, and socio-cultural programme, combining classroom work and community-outreach activities, within an integrated residential environment."

As the Mission Statement indicates, the AEP is somewhat more ambitious than ordinary year-abroad programmes. It brings a cohort of 100 students to spend a full academic year in
Canada on a senate-approved, credit-bearing (i.e., non-ESL) curriculum (see Greenholtz, 2002 for a complete description). In addition to course work, the Japanese undergraduates who come to Canada live in campus residences with Canadian students, engage in language exchanges with Canadian students who are studying Japanese, and participate in a variety of community-outreach activities such as volunteer placements and internship assignments.

However, the AEP shares with all international educational exchanges in the assumption that this type of experience will yield a number of tangible benefits to the participants, and to their respective universities. Indeed, as noted above, the AEP is mandated to contribute to the internationalisation of both the students and the institutions. The Mission Statement is silent on what international understanding should look like in practice, although it sets out the domains in which it should occur; academic, linguistic, and socio-cultural. Therefore, it falls to me as Executive Director to translate intent into specifics, and to assess how well the AEP fulfils its mandate.

Increased language proficiency is invariably first on the list of things that programme participants (according to self reports) most desire from their exchange. We try to help students achieve their linguistic goal both through a
Students gain an understanding of the academic context by following a programme of Senate-approved, credit-bearing courses during their academic year in Canada. These courses require them to perform within North American academic conventions and give them an opportunity to interact with their Canadian peers.

Defining intercultural-understanding goals within the socio-cultural domain has proven to be the most difficult part of the puzzle. Fuzzy notions such as 'better' cross-cultural awareness, sensitivity, or understanding, or 'increased' intercultural competence or effectiveness (see Bradford, Allen, & Beisser, 1998; also Chen, 1997 for a discussion of the differences between competence and effectiveness) usually turn up on self reports as students' most important goal after language proficiency. (The terms cross-cultural and intercultural will be used interchangeably in this paper).

The first set of terms (understanding, awareness, or sensitivity) connotes changes in worldview and appreciation of
other cultures (Bennett, 1986) while the second set (competence, effectiveness) connotes changes within what Ward and Kennedy (1999) call the socio-cultural (behavioural) domain. Despite the similarity in nomenclature, I believe that the term socio-cultural as it appears in the AEP’s Mission Statement refers to more than behaviour. I have taken the socio-cultural to exist within the realm of the socially constructed, and thus have chosen to look for changes in worldview and attitudes towards those different from oneself, rather than simple behavioural adaptations.

Until my thinking about my practice began to evolve in the context of this research, the AEP operated on the unexamined and implicit assumption that international exposure leads to increased awareness or sensitivity in cross-cultural contexts. This was based primarily on the commonly-held notion that exposure, and in the case of the AEP, prolonged, varied, and targeted exposure, is sufficient to act as a catalyst for desired personal growth. To this end, the AEP had consciously built in elements that provide students with opportunities to interact with Canadian society in a variety of ways, both on and off campus. For example, as part of the requirement for LLED 206: Language Field Experience, a compulsory course provided to the AEP by the Faculty of Education, every student
must spend at least 30 hours in a community placement as a volunteer or intern.

Some past examples of community-outreach activities include: teaching swimming to children with disabilities, volunteering at seniors' centres, internships in travel agencies, staffing reception desks at community centres, helping to organise a trade show, working at the SPCA, volunteering with the Down Syndrome Research Society, creating news reports for the community-access cable channel, writing for one of the Japanese community newspapers, working at a food bank, and research assistantships with a quasi-governmental agency.

Unfortunately, there is ample evidence to suggest that exposure alone is insufficient for increasing intercultural sensitivity. In contrast with our optimistic assumptions about the effects of an international exchange experience, Bochner (1986) cites a number of studies in the social psychology literature that indicate that:

increased contact does not necessarily reduce inter-group hostility, and under some conditions actually increases friction and animosity (Bloom, 1971; Mitchell, 1968; Tajfel & Dawson, 1965). Even in culturally mixed (sic) residential settings such as International Houses, where there are explicit pressures to form cross-cultural friendships, studies in the United States, England, and Australia have shown that the various groups prefer the company of their fellow nationals (Bochner, Buker, & McLeod, 1976; Bochner, Hutmik, & Furnham, 1985; Bochner, McLeod, & Lin, 1977; Bochner & Orr, 1979; Furnham &
Alibhai, 1985; Furnham & Bochner, 1982). In many cases, the foreign students had not made a single host-country friend even after a lengthy sojourn (pp. 348-349).

1.3 Defining the Field

It is clear that an empirical investigation of the AEP's impact was required since the quotation above clearly seems to contradict the very raison d'être of exchange programmes. However, personal development in the realm of cross-cultural sensitivity appears not to lend itself very easily to measurement, or even conceptualisation. The trend (see Adler, 1977; Gudykunst & Hammer, 1983, for example) had been to concentrate on observable traits that appeared to comprise intercultural sensitivity.

Bradford et al. (1998) conducted a meta-analysis of intercultural-communication-competence research. While intercultural communication is not the subject of this research, the studies incorporated in the meta-analysis reinforced the impression that the main focus of the field is on observable, performance-based measures or evidence of an ability to perform successfully in cross-cultural environments. Bradford et al. (1998) mention Bennett's (1986) model of development (see below), only as an example of a point of view outside of the mainstream. Most of the literature in the field, as reviewed by Bradford et al. (1998)
is concerned with operationalising overt and explicit manifestations of intercultural effectiveness and appropriateness.

Cross-cultural learning is addressed in other literatures. Business journals, for example, examine the phenomenon from the perspective of maximising the adaptation and effectiveness of employees working in cross-cultural environments. O’Sullivan (1999) looked at the impact of personality traits on trainability in the dynamic cross-cultural skills that constrain the speed of adjustment. Selmer (1999) looked at the relationship between achieving career goals and international adjustment, as measured by subjective well-being. There is also the ever-burgeoning field of cross-cultural psychology which examines the impact of such cultural constructs as individualism and collectivism on cross-cultural assimilation and training (for examples, see Bhawuk, 1998, and Bhawuk & Triandis, 1996).

Ward and Kennedy (1999) have developed a measure of socio-cultural adaptation that taps the behavioural, rather than the attitudinal domain. "[The socio-cultural domain] is related to the ability to ‘fit in’, to acquire culturally appropriate skills and to negotiate interactive aspects of the host environment" (p. 660).
I chose to privilege worldview over the behavioural domain in my practice because behavioural accommodations can be learned without affecting the actor's attitudes towards other cultures. In other words, a person can learn a variety of behavioural adaptations that permit him or her to function in a new culture without understanding why certain behaviours are desirable or to be avoided. Thus, a sojourner can function effectively in a foreign context while continuing to consider the other culture(s) silly, illogical, quaint, or irrelevant. In fact, as Bradford et al. (1998) point out in their discussion of the terms effective and competent, a sojourner can be effective, from the point of view of accomplishing desired goals in business negotiations for example, while acting like the cultural equivalent of a bull in a china shop.

Attitudinal changes result from the conscious, self-reflective processes that Kelly (1963) called construing and reconstruing one's experiences. Although this does not always lead to a more favourable inclination towards the cultures in question, (see discussion of Bennett's DMIS stages below) it does engender a more complex and nuanced experience of life.

Outside the behavioural domain, attempts have been made to measure emotional/affective changes with psychometric tools such as the Zung Self-rating Depression Scale (Zung, 1965) or the Profile of Mood States (McNair, Lorr, & Droppleman, 1971).
These instruments tap the processes of cultural adaptation to a new environment, i.e., culture-shock symptoms such as tension, depression, anger, fatigue, and confusion (Ward & Kennedy, 1999).

Two of the better-known instruments in the cross-culture domain, the Culture Shock Inventory (Reddin, 1994) and the Cross-cultural Adaptability Inventory (Kelly & Meyers, 1992) measure traits thought to be associated with cross-cultural adaptability or sensitivity; also a reflection of the theoretical emphasis in the field alluded to earlier. These traits include flexibility, openness, and emotional strength. A newer entry to the field is Matsumoto et al.'s (2001, as reported in Matsumoto et al., 2003) Intercultural Adjustment Potential Scale (ICAPS). ICAPS is significant in that it was developed for Japanese sojourners and immigrants, instead of U.S. Americans, and recognises that translating an instrument into another language in effect creates a new instrument (more about that below). Its focus, however, is still on adjustment, rather than cross-cultural sensitivity, and on the psychological traits that are posited to facilitate adjustment; emotion regulation, openness, flexibility, and critical thinking.

Chen and Starosta's (2000) Intercultural Sensitivity Scale (ISS) is another instrument that takes a similar
approach. The authors have posited that “individuals must possess six affective elements to be interculturally sensitive: self-esteem, self-monitoring, open-mindedness, empathy, interaction involvement, and suspending judgment” (p. 6). They, therefore, attempted to validate the ISS by comparing correlations of scores on the instrument with existing scales believed to be related; interaction attentiveness, impression rewarding, self-esteem, self-monitoring, and perspective taking. The correlations obtained, while all statistically significant, (ranging from $r=.17$ for the self-esteem scale to $r=.57$ for the perspective-taking scale) are mixed. In statistical circles a correlation of .57 is considered to be outstanding. However, given that a correlation of .57 means that only 27% of the variance is explained by the correlation, compounded by the fact that the relationship among the variables is speculative, I believe the evidence is at this point still too tenuou to account for the construct. Some of these traits make intuitive sense as potential contributors to intercultural sensitivity, but without an overarching theoretical framework to explain the relationship among them, a lot of work remains to be done to demonstrate construct validity for the instrument.
1.4 A Framework for Intercultural Sensitivity

Both Matsumoto et al. (2001) and Chen and Starosta (2000) made significant contributions to the field in exploring the possible constituent elements of intercultural sensitivity. They are complementary to Bennett’s (1986) Developmental Model of Intercultural Sensitivity (DMIS). The DMIS describes the developmental stages through which one progresses on the journey to intercultural sensitivity, but is silent on how one makes that progress, or indeed on what the construct involves except to echo Kelly’s (1963) observation on the necessity to construe and re-construe experience.

As grounded theory based on long years of field observation, the DMIS provides a theoretical framework that sketches what increasing intercultural sensitivity looks like, in a worldview rather than a behavioural sense. That makes it possible to empirically observe and measure the process without necessarily measuring the constituent underlying components.

Bennett’s DMIS, for example, accounts for the increase in friction and animosity, particularly for first-time sojourners; described in the quotation from Bochner (1986) above. The DMIS positions it as the necessary and unavoidable intermediate step from total insularity and lack of awareness of other cultures - the inevitable end result of socialisation.
into one’s own culture (a stage that Bennett calls ‘Denial’) -
towards a broader appreciation of the reality and worth of
other cultural perspectives.

Briefly, the DMIS describes the process of acquiring
intercultural sensitivity in which people move through three
ethnocentric stages to three, more inclusive, ethnoretative
stages.

As mentioned above, the first stage, Denial, is the
default end-product of socialisation into one’s native
culture. In Denial, a person is insensible to the existence of
other cultural norms and worldviews, considering only his or
her own culture or way of thinking to be the natural order of
things. The label Denial refers to the denial of the reality
of other cultural perspectives or values, on any level worth
considering.

When it is no longer possible to ignore the existence of
other cultural points of view, the natural reaction is to
first treat them with hostility and suspicion. This is the
stage that Bennett labels Defence. Defence can take the form
of actively denigrating other cultures or merely asserting the
superiority of one’s own. It is the worldview that underlies
exasperated expatriate statements that begin with, ‘Why can’t
these people learn how to . . .?’ or ‘Why do they do
everything backwards here?’
Further exposure to, and observation of, other cultural points of view can lead to the third ethnocentric stage, Minimization. In Minimization, one views differences in clothing, food, language, or customs as superficial, beneath which people are really pretty much the same. This sounds like a laudable, egalitarian point of view until one remembers that from an ethnocentric worldview, the standard for the sameness of people must be oneself. This worldview still does not allow for other cultures to have an independent reality or validity.

Moving beyond the ethnocentric stages, a person progresses to the DMIS stage of Acceptance, which, as the label implies, brings the realisation that other cultures do have an external reality and are valid on their own terms. Acceptance lays the foundation for Adaptation, both Cognitive and Behavioural, in which other cultures' frames of reference and modes of acting are incorporated into one's own cognitive and behavioural repertoire. In the final stage, Integration, a person can act as a bridge between cultures, having absorbed and incorporated other cultural worldviews. However, he or she operates at the margins of culture, having transcended the confines of a single worldview. The ambivalence contained in the description of Integration, the use of the term 'margins', for example, is intentional. In my understanding, at least, Integration is not as easy to grasp as the other DMIS stages.
and not, as one might expect, simply the promised land at the end of the journey towards perfect intercultural sensitivity. Bennett and Hammer (1998) explain that Integration "describes the effort at integrating disparate aspects of one's identity into a new whole while remaining culturally marginal" (p. 16). Integration can manifest itself as "Constructive marginality (italics in original) ... the form of Integration that refers to the subjective experience of people who are attempting to integrate various cultural identities into larger, more holistic frames of reference" (p. 16) or Encapsulated Marginality, where one is on the fringes, looking in.

Figure 1.1 Continuum of Intercultural Sensitivity

**Continuum of Intercultural Sensitivity**

Milton Bennett - Developmental Model of Intercultural Sensitivity, 1993

- **Defense**
  - Strong defense of one's own world view

- **Minimization**
  - Trivializes differences; focuses on similarities

- **Adaptation**
  - Capable of taking the other's point of view and communicating accordingly

- **Denial**
  - Denies that differences exist

- **Acceptance**
  - Recognizes and Values differences

- **Integration**
  - Values variety of cultures integrates that into behaviour
The DMIS is a social-constructivist model that posits that it takes more than exposure, even of a prolonged nature, to alter one's worldview. As George Kelly (1963, oft cited by Bennett in workshops and lectures) observed:

A person can be witness to a tremendous parade of episodes and yet, if he fails to keep making something out of them . . . he gains little in the way of experience from having been around when they happened. It is not what happens around him that makes a man (sic) experienced; it is the successive construing and reconstruing of what happens, as it happens, that enriches the experience of his life (p. 73).

In other words, it is not enough to be in the vicinity of events as they occur, as Bennett likes to remind us. One must interpret and make sense of those events in order to learn from them, through a process of reflection. In the context of intercultural sensitivity, events must be observed in order to yield data about other cultures, but those data must then be interpreted (reflected upon) to yield information and then knowledge. Conscious reflection upon the feelings and reactions the experience engenders, both towards other cultures and towards one's own, contributes to the growth and evolution of one's worldview.

During the roughly fifteen years I have spent living in Japan, some of it as a 'civilian' and some as a member of Canada's diplomatic corps, and in my continuing involvement with Japanese students, faculty, and staff, I have had ample
opportunity to see in myself and others evidence of the stages that I have come to recognise from Bennett's DMIS model. I have also had the opportunity to learn to engage in the types of self-reflection and self-reflective discussions that have permitted the construing and re-construing of events that Kelly (1963) describes.

A personal anecdote might serve to illustrate the concept. Part of my role as Executive Director is to serve a liaison function between the two universities. In that capacity, I was acting as an intermediary between the Japanese university's administrative staff and a professor visiting from the Canadian university. The professor is fluent in Japanese and is perfectly capable of communicating with staff directly, but the situational context put me at the centre of these exchanges. I was given information concerning deadlines for submitting a syllabus, its length, format, etc. I passed the information on to the professor and she told me she would take care of it. I thought my duties had been fulfilled.

Three weeks later, I received an angry call from the administrative staffer telling me that he hadn't received the syllabus and implying it was my fault. I explained to him that I had given the information to the professor and had received assurances that she would handle it. I suggested to him that
he contact her directly, since she was fluent in Japanese, if there was a problem.

From a North American cultural perspective, I had done as much as could be expected. It would have been improper for me to check up on the professor to confirm she had done as she had promised. I would have been thought to be questioning her sincerity or her competence.

My first reaction as I hung up the phone was, ‘What is that guy’s problem? I’ve done my bit in conveying the information to the professor. The rest is between them.’

Upon further reflection, however, I realised (as I should have at the time of the phone conversation) that from a Japanese perspective the role of intermediary that I had accepted included responsibility for seeing the task through to completion. There is no opting out.

After speaking to the professor, I found out she had been delayed by computer and other problems which, as a North American, she considered beyond her control, and therefore excusable. She had not contacted the administrator to tell him the reason for the delay.

I had to put aside my North American persona and call the administrator back to apologise for failing to fulfil my obligation to see his request through. I also had to call the professor and ask her to fulfil her obligations. I was able to
do this without embarrassment to either of us by making my role as a cultural bridge explicit i.e., that I was acting as way required by Japanese custom, in our shared Japanese context.

Without the reflective stage, I probably would have wasted a considerable amount of psychic energy stewing about the absurdity of being asked to take responsibility for another competent adult’s failure to take action, and learned nothing about my role as a cultural intermediary as a result. This is something I have seen many expatriates do when faced with cross-cultural conundra.

Extrapolating from Kelly (1963), it is clear that the question of whether students on educational exchange programmes become more interculturally sensitive and/or competent is an empirical one. Answering that empirical question is one way in which I can evaluate whether the AEP is providing participants with sufficient exposure and opportunity to provide stimuli to self reflection and, by extension, sufficient opportunity for growth.

1.5 Measuring Intercultural Sensitivity

Bennett and Hammer (1994) attempted to operationalise the DMIS framework by creating the Intercultural Development Inventory (IDI), a psychometric instrument.
The IDI is an English-medium instrument that was developed and validated on English-speaking populations (albeit not all of them U.S. Americans). In IDI training workshops and other fora, the instrument's authors respond to questions about the transferability of the IDI to other than U.S. Americans or English speakers by saying that because the IDI was developed with the participation of subjects from a large number of cultural groups, it can be administered to members of any culture, presumably in any language. They also claim that the IDI has been 'validated' and is therefore ready to pull off the shelf in any context.

In addition to the original reliability and validity analyses by the authors (Bennett & Hammer, 1998), the IDI was the subject of a study by Paige, Jacobs-Cassuto et al. (2003) that sought to "examine the empirical properties of the IDI" (p. 467), specifically to see whether the IDI successfully tracked the underlying DMIS constructs, through a factor analysis. Paige et al., concluded that "the IDI is a reliable measure that has little or no social desirability bias and reasonably, although not exactly, approximates the developmental model of intercultural sensitivity" (p. 467). However, their study was also carried out using U.S. American high school and college students, in a uni-cultural, uni-linguistic context. In order to compensate for the cultural
homogeneity of their subject pool, Paige et al. chose their sample "for variability on a set of personal characteristics that could theoretically be expected to correlate with intercultural sensitivity (e.g., prior international experience, amount of exposure to language education)" (p. 473).

However, the cross-cultural and cross-linguistic transferability and applicability of the DMIS and the IDI remain empirical questions. In validating the ICAPS, Matsumoto et al. (2003) were also cognizant that the transferability of their instrument and the underlying psychological traits was subject to empirical verification. Although they argued that:

... the same psychological skills may (italics added) be necessary for intercultural adjustment of any individuals from any culture as they adjust to a different culture because the psychological skills underlying the process of managing intercultural stress and conflict may (italics added) be the same regardless of culture even though the manifest content of the conflict is culturally specific" (p. 545)

they acknowledge that, "We do not know (italics added) whether or not the ICAPS can predict intercultural adjustment for individuals other than Japanese" (p. 545), and went on to report six studies that investigated the question. I took issue with their casual attitude to the translation of the English original into Japanese, "(A)ll measures were translated into Japanese and their accuracy was verified using
back-translation with no problems" (p. 546) and proposed a rationale for a more comprehensive translation protocol later in this paper. However, they acknowledge, by translating the instrument, that language has the potential to significantly affect the results obtained.

1.6 Linguistic Determinism

In 1929, Edward Sapir first advanced the ideas that he and his student Benjamin Whorf would later propose as the theory of linguistic determinism. In a classic passage, Sapir (1929) argued that:

Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society. It is quite an illusion to imagine that one adjusts to reality essentially without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection. The fact of the matter is that the 'real world' is to a large extent unconsciously built upon the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached. . . We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation (p. 69).

Sapir and Whorf were ahead of their time in understanding that reality is constructed and that one must have categories
for classifying experience, otherwise it (the experience) is meaningless. As Whorf (1956) put it:

We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organised by our minds. . . We cut nature up, organise it into concepts, and ascribe significances as we do (pp. 213-214).

The Sapir-Whorf Hypothesis states, in its strong form, that language determines thought; that language and thought are identical. Linguistic determinism attracts few followers today since there is strong evidence against it, including the possibility of translation between languages. However, the weak version of the Sapir-Whorf Hypothesis, that language strongly influences thought by shaping our categories for classifying experience, is much more widely accepted.

In order to pursue a research agenda based on the DMIS as a theoretical framework, and the IDI as an empirical instrument, my first task, therefore, was to examine the reliability of the IDI in Japanese translation, and the validity of inferences from data gathered from a population that is culturally and linguistically very different from the population on which the aforementioned statistical analyses were performed.
The language above was chosen very deliberately. It is commonly believed that an instrument can be validated, and that once that has been accomplished the instrument is ready for use in any context. However, as Messick (1998) points out, validity is not the property of an instrument. It is a method for determining the appropriateness of inferences drawn from a set of data.

Messick (1998) describes validity thus:

What needs to be valid are the inferences made about score meaning, namely the score interpretation and its action implications for test use. Because value implications both derive from and contribute to score meaning, different value perspectives may lead to different score implications and hence to different validities of interpretation and use for the same scores. This suggests that validity might be indexed to values and perhaps contingent on different facts (or interpretations of facts) surrounding the conditions of use (p. 37).

As we have seen from Sapir and Whorf above, language and culture are significant determinants of values and perspective. It was therefore clearly evident, to me at least, that when an instrument is translated into a different language and administered to a culturally different group of subjects, validity must be examined and established empirically.
CHAPTER TWO

METHODOLOGY

2.1 Introduction

The Intercultural Development Inventory (IDI) is a sixty-item psychometric instrument designed to operationalise the Developmental Model of Intercultural Sensitivity (DMIS) (Bennett, 1986) by identifying test-takers' dominant positioning on a developmental continuum of intercultural sensitivity, and worldview-related issues that remain to be resolved. The IDI is a self-report instrument using a seven-point Likert scale with responses ranging from 'strongly disagree' through 'strongly agree'.

The IDI is a proprietary instrument under copyright to its authors. Anyone who wishes to work with the IDI must attend a qualifying workshop. Certified IDI administrators are required to sign an undertaking not to release the contents of the IDI in whole or in part (individual items). As a result, neither the original instrument nor the Japanese translation appear in this dissertation.

Completing the IDI items requires a relatively high level of English proficiency.

The IDI's authors often state (in IDI qualifying workshops, for example) that the instrument is impervious to
the effects of language and culture because the process of creating it involved input from subjects from a variety of cultural backgrounds. All of that input, however, was in English. In order to make the IDI accessible to my intended research pool, I led a team that translated the IDI into Japanese. That, in turn, raised validity issues; the empirical question of whether inferences from data acquired from the IDI in Japanese translation would be equivalent to those acquired from the original English version.

The Standards for Educational and Psychological Testing (APA, AERA, & NCME, 1985) define validity as "the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (italics added) (p. 9). As the quotation above from Messick (1998) also makes clear, instruments cannot be validated, only the specific inferences from the data (test scores) that they yield. It was important, then, to first establish the validity of the inferences that might be made from the scores from the Japanese translation of the IDI.

Current validity theory, according to Messick (1998), tells us that:

All validity is of one kind, namely, construct validity. Other so-called separate types of validity — whether labelled content validity, criterion-related validity, consequential validity, or whatever — cannot stand alone in validity arguments. Rather, these so-called validity types refer to complementary
forms of evidence to be integrated into an overall judgment of construct validity (p. 37).

In building this evidentiary base, I first looked at the reliability of the scale scores, the extent to which they yielded consistent results, using Cronbach's coefficient alpha and the Intraclass Correlation Coefficient because, as Hubley and Zumbo (1996) remind us, "reliability is a necessary, but not sufficient condition for validity" (p. 208). Following the reliability analysis, I looked at the instrument's construct validity, the extent to which it represented the construct, through an examination of content-validity evidence, criterion-related validity evidence, and a principal components analysis.

2.2 Content Validity

The DMIS is a product of grounded theory that evolved from a long period of observing a phenomenon; in this case, Milton Bennett's long and thoughtful observation of intercultural development. The items that make up the IDI were selected from among statements made in interviews by various people, U.S. Americans and others from various cultural backgrounds living in the United States, about how they viewed and interacted with cultures other than their own. A large number of statements were whittled down, using traditional test-development methods (see the Discussion section below for
a detailed description), to the sixty statements that comprise the first version of the IDI.

Given the grounded nature of the development of the instrument, I thought that the fidelity of the translation to the original would be one piece of evidence for content validity. However, this was not as straightforward as demonstrating linguistic equivalence with translation/back-translation or other measures for assessing translation quality. The process of developing the original IDI used input only from participants (all in the United States although not necessarily U.S. American), who were sufficiently proficient in English to participate in interviews and to contribute statements that could be used verbatim in the item pool. Given the relationship between language and cultural schemata, I knew that I could not take for granted the transferability of the concepts measured by the IDI or inferences from its results, to a culturally and linguistically very different group of subjects.

2.3 Procedures for Translating the IDI

Bennett and Hammer (1998), in reporting their IDI development and validation procedures, stressed the cross-cultural applicability of the instrument (i.e., that a sufficiently diverse cross-section of different languages and
cultures had been represented in their subject pool to eliminate any possible effects on the results from cultural and native-language differences) so I thought it was important to test that assertion by privileging a translation that was faithful to the wording and concepts of the original. However, a number of the translators commented on the 'foreignness' to the Japanese mind of some of the concepts appearing in the instrument. These items were flagged for attention in the statistical analyses (see Results chapter).

There is also a difference between translating and adapting instruments. Ideally, the process of transferring an instrument from one language to another should be a process of adaptation, in which the pith of the theoretical concepts is given precedence.

Test adaptation includes such activities as (1) deciding whether or not a test can measure the same construct in a different language or culture, (2) selecting translators, (3) deciding on appropriate accommodations to be made in preparing a test for use in a second language, and (4) adapting the test and checking its equivalence in the adapted form (Hambleton and Patsula, 1998, p. 155).

Test adaptation also sometimes requires generating a new item where conceptual equivalence is difficult to achieve (Kristjansson, Desrochers, & Zumbo, 2003).

Although the process I followed incorporates many of the points enumerated by Hambleton and Patsula (1988) above, the Japanese-version IDI created for this study was the product of
a conscious process of translation rather than adaptation. For example, I tried to avoid making judgments on conceptual equivalence and accommodations because the cross-cultural transferability of the original instrument and its concepts was the object of empirical verification in this study.

The IDI was first translated into Japanese for an earlier study (Greenholtz, unpublished) that attempted to determine the level of English proficiency, as measured by the Test of English as a Foreign Language (TOEFL), necessary to meaningfully complete the IDI in English. It was the lack of any clear correlation between TOEFL scores and evidence of ability to complete the IDI that convinced me to use a Japanese translation for this study.

Briefly, that original study asked 100 native-Japanese-speaking university undergraduates at various levels of English proficiency each to translate ten of the sixty IDI items into Japanese. They were also asked to state their level of confidence with their translation. That was an effort to indirectly identify concepts in the instrument that might pose extra-linguistic (conceptual) difficulties for Japanese speakers. The translations were rated and statistically analysed with TOEFL scores as the independent variable. As noted earlier, no clear correlation with TOEFL scores emerged with predictive value from those data.
In order to rate the undergraduates' responses for that study, it had been necessary to create a Japanese IDI 'master' translation.

Four Japanese-English bilinguals (native Japanese speakers) were each asked to independently translate all sixty items into Japanese. They were judged to be bilingual because they had each completed post-graduate work in the United States (one at the doctoral level). One of the translators, who had completed an MBA at an American university, was working in the United States for a large Japanese multinational corporation. Two others, one of whom was also a certified IDI administrator, were experienced English teachers working in Japan. The fourth translator had completed a PhD from Columbia University in second-language acquisition, and was also living in the United States.

My qualifications for leading the translation team were as follows. I am a native English speaker who has spent over half of my adult life in Japan. I passed the highest level of the Japanese Language Proficiency Test (Nihongo Noryoku Shiken, Japan Foundation) in 1984 and have done a large number of Japanese-to-English translations and published an English-to-Japanese translation of My Friend David (Edwards and Dawson, 1983) entitled Mai Frendo Deibido (Greenholtz and Morita, 1988). I was certified as a Japanese-English court
interpreter by the government of Ontario in 1991 and serve as a consecutive interpreter in a variety of contexts. I am also a certified administrator of the IDI.

I selected a final version from the pool of items (the four independent translations) and this version was given to two raters, native-Japanese-speaking doctoral candidates at a Canadian university, to rate the undergraduates' translations. The rating exercise required the raters to consider the items much more deeply than simply comparing them to the English originals or performing a back translation would have, because they had to justify the score they had assigned to each of the translations.

Following the rating phase, I discussed each item with the two raters for fidelity to the language and concepts of the original, within the constraints of natural Japanese. Several changes to wording of items were suggested and many of those changes were incorporated into what became the Japanese-language version used in the present study.

This procedure far exceeded the 'translate/back-translate' procedure that had long been considered the 'gold standard' for translating material into other languages. The method employed exceeded even the 'new gold standard' described in Kristjansson, Desrochers, and Zumbo (2003) who (citing Behling & Law, 2000; Hambleton & Patsula, 1998) point
out, "direct translation and back translation can deal with literal meaning only", and "(B)ack translation cannot detect differences in conceptual understanding of the question, and so cannot ensure psychological equivalence of the items in a scale or questionnaire" (p. 135). Although my intention was to use as literal a translation as possible, I also needed to be cognizant of conceptual inconsistencies or difficulties posed to Japanese speakers by the IDI because I wasn't interested only in producing a translation. I was also examining whether there was a sufficiently strong evidentiary validity argument for using the translation. As part of that process, the conceptual nuances helped me to make sense of the principal components analysis.

The translation process is summarised in Table 2.1, below.

Despite the multi-step process followed in the translation, a few clerical errors slipped in that were discovered only after the test was administered. One was a typographical error in which the transformation from Roman characters to Japanese text was only partially completed, leaving an English letter in the middle of a Japanese word. None of the many participants questioned after this error was discovered thought that the typographical error obscured the item's meaning in any way.
In two of the items a modifier, 'some', was left out, in another 'often' was omitted, and in a fourth, 'nearly'. These items were all flagged for attention during the analysis (see Results).

Finally, there was an ambiguous choice of characters (between two characters with the same reading, and similar meaning, that are, however, followed by different verb complements). This item was also flagged for attention during the analysis.

These errors came to light when my research assistant (RA), a Japanese-English bilingual undergraduate student,
completed both the English and Japanese versions at one sitting. We compared her scores on an item-by-item basis and where they were different, tried to determine why. Some of the differences were due to items that had concepts that did not transfer well to a Japanese cultural framework (as flagged by the translators, see discussion above), but were conceptually available to her in the original English version. The others revealed the errors discussed above.

Based on this experience, I would recommend that this technique (having a bilingual complete both versions of the instruments at the same sitting and comparing the results on an item-by-item basis) be added to the protocol for translating/adapting instruments. Although this should have been done during the translation phase, the information about these items was still useful in the analysis.

2.4 Assessing the Adequacy of Construct Representation

2.4.1 Participants

Four hundred students completed the Japanese version of the IDI, over a three-year period. All of the IDIs were completed before the students began their exchange experience. The students were all undergraduates at a large private university in Japan. Two hundred eighty-seven (287) were female and 113 were male. Three hundred and seventy-one (371)
of the students were in second year, twenty-seven (27) in third year, and two (2) were fourth-year students. Participants ranged in age from nineteen (19) to forty-eight (48) years. Ninety-five (95) percent of the participants were between nineteen and twenty-one years of age.

These IDIs provided the data for assessing the construct validity of the Japanese version.

2.4.2 Principal Components Analysis

The 400 IDIs were analysed using principal components analysis, first with Direct Quartimin rotation (see Zumbo & Taylor, 1993) to determine whether there were any correlations among the factors. Once it was confirmed that the inter-factor correlations were small, Varimax rotation was used to determine whether the factor structure of the Japanese version matched that of the English original, for this population. The results were further confirmed with a factor analysis (Maximum Likelihood with Varimax rotation).

A hybrid exploratory/confirmatory approach was used for the principal components analysis. It was not entirely exploratory because the results of the analyses of the original English IDI guided the determination of the general parameters for this analysis. However, since it wasn't a certainty that the Japanese version would be identical to the original, the analysis was not wholly confirmatory either. In
addition, following Kelloway's (1998) guideline that confirmatory factor analysis requires a 10:1 subject to item ratio, the data set, with only 400 subjects for a sixty-item instrument was not large enough for a purely confirmatory factor analysis.

Principal components analyses were performed for four- through seven-factor models to find the best fit. Knowing that there ought to be in the neighbourhood of six factors permitted the selection of initial Eigenvalues greater than 1.5 (rather than the more traditional rule-of-thumb of 1.0), which greatly simplified the analysis.

The four-, five-, six-, and seven-factor models were analysed and interpreted using the rotated factor matrix, the scree plot, and nonredundant residuals with absolute values greater than 0.05, to determine which made the most sense both statistically and interpretively.
CHAPTER THREE
RESULTS

Following Zumbo and Taylor (1993) an analysis was first performed with Direct Quartimin rotation (Direct Oblimin with Delta set at zero) to determine whether there was any correlation between the factors (see Table 3.1). Factors with initial Eigenvalues of 1.0 or greater were retained. This yielded fourteen factors that accounted for 57.56% of the variance. The analysis and validation of the original IDI (Bennett and Hammer, 1998) led me to expect somewhere in the neighbourhood of six factors so factors with initial Eigenvalues of greater than 1.5 were retained instead. This brought the number of factors down to seven, accounting for 42.296% of the variance.

Table 3.1 Component Correlation Matrix

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.093</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-.042</td>
<td>-.088</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.187</td>
<td>.224</td>
<td>-.116</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-.205</td>
<td>-.015</td>
<td>-.052</td>
<td>-.319</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-.087</td>
<td>.039</td>
<td>-.058</td>
<td>-.087</td>
<td>.145</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.167</td>
<td>-.010</td>
<td>-.077</td>
<td>.022</td>
<td>-.012</td>
<td>.011</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Component Analysis.
Rotation Method: Oblim in with Kaiser Normalization.
The Kaiser-Meyer-Olkin measure of sampling adequacy was used (Kaiser, 1970 as cited in Zumbo & Taylor, 1993) yielding a respectable value of .85 (Table 3.2).

Table 3.2 KMO Measure of Sampling Adequacy and Bartlett's Test of Sphericity

<table>
<thead>
<tr>
<th>Kaiser-Meyer-Olkin Measure of Sampling Adequacy.</th>
<th>.850</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett's Test of Sphericity</td>
<td></td>
</tr>
<tr>
<td>Approx. Chi-Square</td>
<td>8031.29</td>
</tr>
<tr>
<td>df</td>
<td>1770</td>
</tr>
<tr>
<td>Sig.</td>
<td>.000</td>
</tr>
</tbody>
</table>

Once it was determined that the correlations among the factors were very small, four-, five-, six-, and seven-factor solutions were tried with Varimax rotation.

This decision was also based on foreknowledge from the validation process for the English-language (original version) of the IDI (Bennett & Hammer, 1998) and a second analysis by Paige et al. (2003) indicating that the instrument measured six discrete factors.

The scree plot (Figure 3.1), rotated factor matrix (Table 3.3), and nonredundant residuals were all considered in determining which model yielded the best interpretation of the data. The scree plot identified five fairly clear factors. It is always difficult to decide where the scree begins and
factors six and seven were in a position that made it difficult to judge whether they were significant or not. The reason for this ambiguity became clearer in the analysis of the rotated factor matrix. The progression from four- to seven-factor solution models brought the nonredundant residuals with absolute values greater than .05 down from 30% to 20%.

Figure 3.1 Scree Plot

Scree Plot

![Scree Plot](image)

Component Number
<table>
<thead>
<tr>
<th>Item</th>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>24-(mo) Cognitive Adaptation-Bridge Builder</td>
<td>.673</td>
<td>.024</td>
<td>-.025</td>
<td>.257</td>
<td>-.041</td>
<td>.015</td>
<td>-.015</td>
<td></td>
</tr>
<tr>
<td>18-Cognitive Adaptation-Bridge Builder</td>
<td>.633</td>
<td>.030</td>
<td>.017</td>
<td>.085</td>
<td>-.062</td>
<td>.048</td>
<td>.033</td>
<td></td>
</tr>
<tr>
<td>53-Cognitive Adaptation-Bridge Builder</td>
<td>.606</td>
<td>-.104</td>
<td>-.019</td>
<td>-.140</td>
<td>-.089</td>
<td>-.048</td>
<td>-.027</td>
<td></td>
</tr>
<tr>
<td>50-Behavioural Adaptation-Behavioural Shift</td>
<td>.597</td>
<td>.044</td>
<td>.226</td>
<td>.134</td>
<td>-.012</td>
<td>.088</td>
<td>-.016</td>
<td></td>
</tr>
<tr>
<td>60-Cognitive Adaptation-Bridge Builder</td>
<td>.597</td>
<td>.061</td>
<td>.225</td>
<td>.020</td>
<td>.047</td>
<td>-.075</td>
<td>.087</td>
<td></td>
</tr>
<tr>
<td>25-Cognitive Adaptation-Frame Shifting</td>
<td>.589</td>
<td>.021</td>
<td>.185</td>
<td>-.006</td>
<td>.058</td>
<td>.011</td>
<td>.015</td>
<td></td>
</tr>
<tr>
<td>52-Cognitive Adaptation-Frame Shifting</td>
<td>.573</td>
<td>-.079</td>
<td>-.086</td>
<td>-.058</td>
<td>-.071</td>
<td>-.152</td>
<td>.036</td>
<td></td>
</tr>
<tr>
<td>46-Cognitive Adaptation-Frame Shifting</td>
<td>.547</td>
<td>-.066</td>
<td>-.025</td>
<td>-.023</td>
<td>.020</td>
<td>-.048</td>
<td>-.031</td>
<td></td>
</tr>
<tr>
<td>37-Denial-Disinterest</td>
<td>.536</td>
<td>-.006</td>
<td>.014</td>
<td>.286</td>
<td>.039</td>
<td>-.005</td>
<td>.020</td>
<td></td>
</tr>
<tr>
<td>36-Behavioural Adaptation-Cultural Complexity</td>
<td>.499</td>
<td>-.092</td>
<td>-.038</td>
<td>.085</td>
<td>.032</td>
<td>-.045</td>
<td>-.062</td>
<td></td>
</tr>
<tr>
<td>26-Behavioural Adaptation-Cultural Complexity</td>
<td>.462</td>
<td>.050</td>
<td>.083</td>
<td>.092</td>
<td>-.031</td>
<td>.035</td>
<td>.076</td>
<td></td>
</tr>
<tr>
<td>13-Behavioural Adaptation-Behavioural Shift</td>
<td>.442</td>
<td>.093</td>
<td>.069</td>
<td>-.075</td>
<td>-.051</td>
<td>-.024</td>
<td>.023</td>
<td></td>
</tr>
<tr>
<td>35-Behavioural Adaptation-Cultural Complexity</td>
<td>.440</td>
<td>.041</td>
<td>.103</td>
<td>.063</td>
<td>-.096</td>
<td>.008</td>
<td>.049</td>
<td></td>
</tr>
<tr>
<td>3-Cognitive Adaptation-Frame Shifting</td>
<td>.405</td>
<td>.019</td>
<td>.070</td>
<td>-.006</td>
<td>.091</td>
<td>.014</td>
<td>.005</td>
<td></td>
</tr>
<tr>
<td>9-(nearly) Behavioural Adaptation-Cultural Complexity</td>
<td>.385</td>
<td>-.093</td>
<td>.087</td>
<td>.247</td>
<td>-.105</td>
<td>.048</td>
<td>-.070</td>
<td></td>
</tr>
<tr>
<td>7-Behavioural Adaptation-Behavioural Shift</td>
<td>.365</td>
<td>.038</td>
<td>.057</td>
<td>.066</td>
<td>.131</td>
<td>-.108</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td>58-Behavioural Adaptation-Behavioural Shift</td>
<td>.331</td>
<td>-.041</td>
<td>.254</td>
<td>-.164</td>
<td>-.067</td>
<td>-.139</td>
<td>.047</td>
<td></td>
</tr>
<tr>
<td>17-Acceptance-Enjoying Difference</td>
<td>.322</td>
<td>.126</td>
<td>.212</td>
<td>.165</td>
<td>.024</td>
<td>.094</td>
<td>-.036</td>
<td></td>
</tr>
<tr>
<td>54-Behavioural Adaptation-Behavioural Shift</td>
<td>.268</td>
<td>-.054</td>
<td>.154</td>
<td>-.222</td>
<td>-.108</td>
<td>-.100</td>
<td>.035</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>Factor 1</td>
<td>Factor 2</td>
<td>Factor 3</td>
<td>Factor 4</td>
<td>Factor 5</td>
<td>Factor 6</td>
<td>Factor 7</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>20-Defence-Denigration</td>
<td>.015</td>
<td>.732</td>
<td>.135</td>
<td>-.005</td>
<td>-.045</td>
<td>.115</td>
<td>.015</td>
<td></td>
</tr>
<tr>
<td>28-Defence-Denigration</td>
<td>.056</td>
<td>.708</td>
<td>.174</td>
<td>.051</td>
<td>.063</td>
<td>.052</td>
<td>.003</td>
<td></td>
</tr>
<tr>
<td>41-Defence-Denigration</td>
<td>.054</td>
<td>.681</td>
<td>.237</td>
<td>.169</td>
<td>.035</td>
<td>-.069</td>
<td>.033</td>
<td></td>
</tr>
<tr>
<td>16-Defence-Superiority</td>
<td>-.004</td>
<td>.678</td>
<td>.125</td>
<td>.042</td>
<td>-.006</td>
<td>.100</td>
<td>.036</td>
<td></td>
</tr>
<tr>
<td>55-Defence-Superiority</td>
<td>-.058</td>
<td>.653</td>
<td>.214</td>
<td>.216</td>
<td>-.046</td>
<td>-.031</td>
<td>.036</td>
<td></td>
</tr>
<tr>
<td>11-Defence-Superiority</td>
<td>.010</td>
<td>.644</td>
<td>.042</td>
<td>.013</td>
<td>.077</td>
<td>.095</td>
<td>-.015</td>
<td></td>
</tr>
<tr>
<td>39-Defence-Superiority</td>
<td>-.023</td>
<td>.582</td>
<td>.228</td>
<td>.095</td>
<td>.196</td>
<td>.018</td>
<td>-.073</td>
<td></td>
</tr>
<tr>
<td>10-Defence-Denigration</td>
<td>-.034</td>
<td>.517</td>
<td>.123</td>
<td>.098</td>
<td>-.011</td>
<td>.087</td>
<td>.035</td>
<td></td>
</tr>
<tr>
<td>44-Defence-Denigration</td>
<td>.049</td>
<td>.514</td>
<td>.137</td>
<td>.377</td>
<td>.045</td>
<td>-.021</td>
<td>.053</td>
<td></td>
</tr>
<tr>
<td>56-Defence-Superiority</td>
<td>-.108</td>
<td>.458</td>
<td>.222</td>
<td>.035</td>
<td>.202</td>
<td>.060</td>
<td>-.034</td>
<td></td>
</tr>
<tr>
<td>14-Denial-Avoidance Separation</td>
<td>.030</td>
<td>.240</td>
<td>.153</td>
<td>.141</td>
<td>-.092</td>
<td>-.024</td>
<td>-.098</td>
<td></td>
</tr>
<tr>
<td>31-Cognitive Adaptation-Multiple Perspective</td>
<td>.057</td>
<td>.070</td>
<td>.566</td>
<td>.050</td>
<td>.009</td>
<td>-.068</td>
<td>-.035</td>
<td></td>
</tr>
<tr>
<td>29-Acceptance-Enjoying Difference</td>
<td>.019</td>
<td>.119</td>
<td>.564</td>
<td>.196</td>
<td>-.106</td>
<td>-.022</td>
<td>.103</td>
<td></td>
</tr>
<tr>
<td>32-Acceptance-Describing Difference</td>
<td>.025</td>
<td>.090</td>
<td>.524</td>
<td>-.052</td>
<td>.085</td>
<td>.047</td>
<td>.132</td>
<td></td>
</tr>
<tr>
<td>57-Denial-Disinterest</td>
<td>.117</td>
<td>.262</td>
<td>.509</td>
<td>.147</td>
<td>.017</td>
<td>.012</td>
<td>-.058</td>
<td></td>
</tr>
<tr>
<td>27-Acceptance-Learning Difference</td>
<td>.343</td>
<td>.112</td>
<td>.470</td>
<td>.190</td>
<td>-.045</td>
<td>-.014</td>
<td>-.032</td>
<td></td>
</tr>
<tr>
<td>34-Minimisation-Superficial Differences</td>
<td>.001</td>
<td>-.139</td>
<td>-.460</td>
<td>-.143</td>
<td>.419</td>
<td>-.046</td>
<td>.067</td>
<td></td>
</tr>
<tr>
<td>21-Acceptance-Value Relativity</td>
<td>.069</td>
<td>.183</td>
<td>.448</td>
<td>.089</td>
<td>.008</td>
<td>.073</td>
<td>.047</td>
<td></td>
</tr>
<tr>
<td>15-Denial-Disinterest</td>
<td>.108</td>
<td>.233</td>
<td>.423</td>
<td>.209</td>
<td>-.066</td>
<td>.004</td>
<td>-.054</td>
<td></td>
</tr>
<tr>
<td>19-Acceptance-Learning Difference</td>
<td>.105</td>
<td>.099</td>
<td>.391</td>
<td>.071</td>
<td>-.068</td>
<td>.066</td>
<td>-.004</td>
<td></td>
</tr>
<tr>
<td>33-Acceptance-Describing Difference</td>
<td>.040</td>
<td>.171</td>
<td>.377</td>
<td>.077</td>
<td>-.050</td>
<td>-.020</td>
<td>-.045</td>
<td></td>
</tr>
<tr>
<td>1-Denial-Disinterest</td>
<td>.059</td>
<td>.167</td>
<td>.359</td>
<td>.143</td>
<td>-.027</td>
<td>-.047</td>
<td>-.160</td>
<td></td>
</tr>
<tr>
<td>47-Acceptance-Enjoying Difference</td>
<td>.073</td>
<td>.140</td>
<td>.323</td>
<td>.075</td>
<td>.028</td>
<td>.098</td>
<td>.052</td>
<td></td>
</tr>
<tr>
<td>45-Cognitive Adaptation-Multiple Perspective</td>
<td>.245</td>
<td>.028</td>
<td>.283</td>
<td>-.035</td>
<td>-.119</td>
<td>.055</td>
<td>.128</td>
<td></td>
</tr>
<tr>
<td>42-Behavioural Adaptation-Behavioural Shift</td>
<td>.201</td>
<td>.052</td>
<td>.224</td>
<td>-.030</td>
<td>-.160</td>
<td>.140</td>
<td>.057</td>
<td></td>
</tr>
<tr>
<td>Item</td>
<td>Factor</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>38-Denial-Avoidance Separation</td>
<td></td>
<td>.096</td>
<td>.222</td>
<td>.277</td>
<td>.668</td>
<td>-.008</td>
<td>-.019</td>
<td>.028</td>
</tr>
<tr>
<td>40-Denial-Avoidance Separation</td>
<td></td>
<td>.151</td>
<td>.216</td>
<td>.264</td>
<td>.655</td>
<td>.041</td>
<td>.046</td>
<td>-.091</td>
</tr>
<tr>
<td>43-Denial-Avoidance Separation</td>
<td></td>
<td>.081</td>
<td>.333</td>
<td>.199</td>
<td>.640</td>
<td>.054</td>
<td>.027</td>
<td>-.033</td>
</tr>
<tr>
<td>49-Denial-Avoidance Separation</td>
<td></td>
<td>.186</td>
<td>.241</td>
<td>.339</td>
<td>.566</td>
<td>-.068</td>
<td>.015</td>
<td>-.030</td>
</tr>
<tr>
<td>30-Denial-Avoidance Separation</td>
<td></td>
<td>.103</td>
<td>.298</td>
<td>.266</td>
<td>.364</td>
<td>-.065</td>
<td>.052</td>
<td>-.007</td>
</tr>
<tr>
<td>6-Minimisation-Human Similarities</td>
<td></td>
<td>-.111</td>
<td>.129</td>
<td>-.073</td>
<td>-.167</td>
<td>.126</td>
<td>.021</td>
<td>.115</td>
</tr>
<tr>
<td>22-Minimisation-Human Similarities</td>
<td></td>
<td>-.004</td>
<td>-.005</td>
<td>-.071</td>
<td>-.056</td>
<td>.670</td>
<td>-.010</td>
<td>.173</td>
</tr>
<tr>
<td>4-Minimisation-Superficial Differences</td>
<td></td>
<td>.023</td>
<td>.015</td>
<td>-.091</td>
<td>.045</td>
<td>.564</td>
<td>.025</td>
<td>-.008</td>
</tr>
<tr>
<td>51-Minimisation-Human Similarities</td>
<td></td>
<td>-.100</td>
<td>.187</td>
<td>.038</td>
<td>-.152</td>
<td>.498</td>
<td>.206</td>
<td>-.053</td>
</tr>
<tr>
<td>8-Minimisation-Superficial Differences</td>
<td></td>
<td>.032</td>
<td>.048</td>
<td>.033</td>
<td>.054</td>
<td>.477</td>
<td>.083</td>
<td>-.046</td>
</tr>
<tr>
<td>5-Minimisation-Universal Values</td>
<td></td>
<td>-.022</td>
<td>.033</td>
<td>-.068</td>
<td>.022</td>
<td>.429</td>
<td>.091</td>
<td>.097</td>
</tr>
<tr>
<td>59-(sb) Minimisation-Universal Values</td>
<td></td>
<td>-.075</td>
<td>.112</td>
<td>.054</td>
<td>.057</td>
<td>.101</td>
<td>.834</td>
<td>-.041</td>
</tr>
<tr>
<td>12-(sb) Minimisation-Universal Values</td>
<td></td>
<td>-.087</td>
<td>.149</td>
<td>.060</td>
<td>.023</td>
<td>.168</td>
<td>.695</td>
<td>.065</td>
</tr>
<tr>
<td>23-(sb) Minimisation-Universal Values</td>
<td></td>
<td>-.052</td>
<td>.056</td>
<td>.040</td>
<td>.000</td>
<td>.194</td>
<td>.305</td>
<td>.156</td>
</tr>
<tr>
<td>2-(ms) Acceptance-Value Relativity</td>
<td></td>
<td>.034</td>
<td>-.019</td>
<td>.039</td>
<td>-.028</td>
<td>.124</td>
<td>.092</td>
<td>.851</td>
</tr>
<tr>
<td>48-(ms) Acceptance-Value Relativity</td>
<td></td>
<td>.123</td>
<td>.011</td>
<td>.038</td>
<td>-.054</td>
<td>.053</td>
<td>.024</td>
<td>.624</td>
</tr>
<tr>
<td>% of Variance</td>
<td></td>
<td>14.923</td>
<td>9.436</td>
<td>5.132</td>
<td>3.757</td>
<td>3.499</td>
<td>2.880</td>
<td>2.668</td>
</tr>
<tr>
<td>Cumulative % of Variance</td>
<td></td>
<td>14.923</td>
<td>24.359</td>
<td>29.491</td>
<td>33.248</td>
<td>36.748</td>
<td>39.628</td>
<td>42.296</td>
</tr>
<tr>
<td>Eigenvalues</td>
<td></td>
<td>8.954</td>
<td>5.661</td>
<td>3.079</td>
<td>2.254</td>
<td>2.100</td>
<td>1.728</td>
<td>1.601</td>
</tr>
</tbody>
</table>

In analysing the rotated factor matrix (Table 3.3), traditional rule-of-thumb indicators for factor loadings (.45 or more is usually considered a strong loading, loadings
between .30 and .45 are moderate, and loadings less than .30 are usually considered too weak) were used.

The transition zones between factors contained items that loaded at nearly .30, but also loaded almost as highly across the matrix with other factors. These were reduced in number with successive iterations from four to seven factors. Finally, given that these items did not belong to the factor indicated by the rest of the items they were notionally excluded from the overall judgment regarding what the factor mainly indicated. This process was greatly aided by choosing the SPSS option to have the factor loadings listed by size.

The seven-factor solution was the most satisfactory. It yielded five fairly coherent factors, the first of which appeared to be comprised of two subscales, and two additional factors that reflected issues raised in the translation process (see Discussion).

3.1 The Seven-Factor Model

In the seven-factor model, the first factor corresponded to the IDI's Adaptation dimension, although it came out as one factor rather than the two, Cognitive Adaptation and Behavioural Adaptation, that the original IDI contains. Having said that, while they could not be said to be separate factors, it was generally true that Cognitive Adaptation items
were clustered together with higher loadings than the Behavioural Adaptation items, so a convincing argument can be easily made for two subscales within a main factor. This corresponds with the IDI's structure. Overall, fifteen of the twenty Adaptation items loaded on this factor.

The second factor was clearly the IDI's Defence dimension, with all ten items loading cleanly and exclusively (factor loadings ranging from .754 to .499) on this factor.

The third factor was slightly less clean. It contained seven of ten items from the IDI's Acceptance dimension, but the highest loading item on the factor (.649) was an orphaned Cognitive Adaptation item and there were also three items from the Denial dimension in the mix. Overall, the factor was dominated by the Acceptance dimension, with loadings for Acceptance items ranging from .612 to .380. Given that Acceptance and Adaptation are adjacent dimensions on the scale, the fact that an Adaptation item would load highly on the Acceptance factor is not beyond comprehension.

In addressing the relatively lower internal consistency of the Acceptance dimension, Paige et al. (2003) noted the "complexity and multidimensionality of this first ethnorelative stage, which in operational terms, is reflected in a larger variety of statement types used to measure this construct" (p. 483). Four subscales, describing difference,
enjoying difference, learning difference and value relativity all co-exist within the ten-item scope of this dimension. An alternative to Paige et al.'s interpretation will be presented in the Discussion chapter.

The fourth factor (following a transition area containing two orphaned Adaptation items) contained five of the items from the Denial scale. This is only half of the Denial items, but they clearly dominated the factor with loadings ranging from .690 to .381.

The fifth factor corresponded to the Minimization dimension of the IDI, containing six of the ten Minimization items, although the last item loaded almost as strongly on the Acceptance factor. This is somewhat surprising because although Minimization and Acceptance are adjacent dimensions, they are on opposite sides of the Ethnocentric – Ethnorelative divide.

The other two factors did not fit the IDI structure, but made sense in the context of conceptual difficulties that the Japanese speakers brought to light during the translation process.

The sixth factor was a neat cluster of the three items on the Universal Values subscale of the Minimisation dimension that contained references to all humans being 'children of a spiritual being'. All of the native-Japanese-speaking
translators found this to be a truly foreign concept. One feature of Japanese that makes it so welcoming of foreign loan words is a syllabic script (katakana) used almost exclusively for rendering those loan words into the nearest Japanese phonetic equivalent. (How the meanings of those words diffuse through the Japanese lexical awareness should be the topic of what I am sure would be a fascinating socio-linguistic study). Thus it was possible to have the phrases, rendered phonetically in *katakana*, marked with an asterisk leading to a glossary at the bottom of the form listing alternative dictionary definitions. Although three of sixty items is not that significant in the larger scheme of things, and do not by themselves compromise the utility of the instrument for a Japanese-speaking population, it was cause for some reflection within the context of a multi-dimensional search for validity evidence. This problem will also be dealt with at greater length in the Discussion section.

The seventh factor contained two items that had been flagged from the translation for two reasons. One was the omission of the modifier 'some' from the Japanese rendition. The second was that they were actually the same item. The original English items are nearly identical except that one asks whether it is appropriate for cultures to have different conceptions of 'right and wrong' and other, different
conceptions of ‘good and bad’. I hadn’t noticed that the translators had used the same term in both instances (and neither, apparently, had they) until my RA pointed it out when we were examining why her scores for those items differed between the English and Japanese versions. The difference (in her case) was due to the omission of the word ‘some’ from the translations, (she also hadn’t noticed that the same Japanese translation had been used for both items) but that discussion brought to the light the fact that the items were identical.

Whether the clustering on a clearly distinct seventh factor (those items had the highest loading in the entire matrix of .798 and .768 respectively) was due to the omission of ‘some’ (unlikely, in my view, since the omission of the word ‘often’ from item 24 did not prevent it from having the highest loading on the Adaptation factor) or the fact that the items were the same, (that again being no explanation as to why they should form a distinct factor) or to the fact that the concept of ‘good and bad’ somehow represented a discrete factor to Japanese participants, is unclear. Further discussion with other Japanese-English bilinguals has revealed that it is difficult for them to provide clearly distinct lexical alternatives (‘right’ and ‘wrong’ usually comes out first as some variation of ‘correct’ and ‘incorrect’, i.e., as adjectives rather than nouns). This is not to say that lexical
equivalents cannot be found, only that they must be actively searched for, or perhaps more context provided in the sentence. One translator suggested that without a complement the concept would prove significant syntactic difficulties in Japanese.

Because the true nature of the seventh factor was compromised in a variety of ways and the sixth factor is explainable in terms of the foreignness of the concept of 'child of a spiritual being', I judged that the Japanese version of the IDI could be characterised as having five clear factors, with one, Adaptation, comprising two sub-factors.

For confirmation I ran a factor analysis (Maximum Likelihood with Varimax rotation) on the seven-factor model. The results of this analysis were essentially identical with the principal components analysis (some items within the factors were in a different order) except that the nonredundant residuals with an absolute value greater than 0.05 were further reduced to 15%.

Another piece of the evidentiary puzzle for validity is reliability. To demonstrate the reliability of the Japanese IDI Cronbach's coefficient alpha was computed. The results (see Table 3.4) yielded a coefficient alpha of 0.854.
Table 3.4 Cronbach's Coefficient Alpha

<table>
<thead>
<tr>
<th>Reliability Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of Cases = 400.0</td>
</tr>
<tr>
<td>N of Items = 60</td>
</tr>
<tr>
<td>Alpha = .8540</td>
</tr>
</tbody>
</table>

These analyses appeared to establish the essential equivalence of the Japanese translation of the IDI with the original English version. They reflect quite closely results reported by Bennett and Hammer (1998) and Paige et al. (2003). Statistical analyses can be accepted uncritically at face value; factor loadings above .45 or correlations of .8 or greater trotted out to 'demonstrate' validity, but the validation process really requires active interpretation on the part of the researcher, whatever story the numbers might appear to tell. Looking beneath the numerical results, the data raise questions about whether the IDI (and by extension, the DMIS that it is meant to operationalise) is 'culture free' or applicable to any cultural context. This will be further explored in the Discussion chapter.
4.1 Validity Revisited

This research was not initially intended to be about validity issues or the examination of validity in a cross-cultural context. When I learned about the IDI, I had been excited to find out that a 'reliable and valid' instrument existed to operationalise the DMIS, and was keen to use the IDI as a tool for gaining some empirical insights into the effectiveness of the academic exchange programme I was administering. Like most practitioners, I was mostly unaware that there might be reasons to be cautious about using it in my particular research context. However, the English-language proficiency of my subject pool was not sufficiently high to use the IDI in its original form and that necessitated a translation into Japanese. I knew enough to suspect that changing the language of the instrument might affect its reliability and validity, although I was still not very conversant in validity issues. I had hoped to quickly 'validate' the Japanese version of the IDI in order to use the instrument in the next stage of my research, not yet knowing that validity was a property of inferences drawn from data, not of the instrument itself. Further investigation into the
field of validity made it quickly clear that differences in language and culture posed potentially critical validity issues.

This is particularly important for other practitioners in academic exchanges or others involved in education in cross-cultural contexts who may not be conversant with validity theory. A case in point was a recent study by Westrick (2004) who used the IDI in a quantitative analysis of the effect of service learning on the intercultural sensitivity of high school students at an international school in Hong Kong. Westrick’s (2004) rationale for using the IDI was that:

Psychometric analysis shows that ‘the IDI is a highly reliable measure which has little or no social desirability bias and also reasonably approximates the developmental model of intercultural sensitivity (Bennett, 1986, 1993) upon which it is based’ (Paige et al., 1999). Now that there is a reliable instrument to measure these theoretical stages (Hammer et al., 2002) there is a potential to gain insights into international school students and the different programs and strategies that can increase their intercultural sensitivity (italics added) (p. 282).

This rationale, which represented my own starting point in embarking on this research, now rings serious alarm bells for me. First, it mentioned only the reliability of the instrument, and not its validity. It accepted without question that the instrument would be suitable for use in the context of high school students, from a variety of cultural
backgrounds, where "over a third, 38.7 percent of the students in the sample claim nationalities in an Asian country" (Westrick, 2004, p. 286).

In her conclusion, Westrick (2004) went on to state that, "(A)s a statistically valid and reliable instrument, the IDI has been shown to be a valuable tool for evaluating students' level of intercultural sensitivity" (italics added) (p. 296). However, she made no mention of having considered potential validity issues in any way. This was disturbing for two reasons. The first was the uncritical perpetuation of the notion that an instrument can be validated. The second involved an aspect of validity theory that lies beyond what have traditionally been regarded as the elements of construct validity. Messick (1998) calls it consequential validity, "the relation between the evidential and consequential bases of validity" (p. 38) (about which more will be said below).

The DMIS describes the process of increasing one's intercultural sensitivity as a constructivist one. Westrick (2004) summarised the requirements as follows:

Finally, Bennett articulates three assumptions about the development of intercultural sensitivity, all of which have particular relevance to the experience of students involved in service-learning.
1. The phenomenology of difference is the key to intercultural sensitivity.
2. The construing of difference necessary for intercultural sensitivity is that of ethnorelativism, whereby different cultures are perceived as variable and viable constructions of reality.
3. *Ethical choices* can and must be made for intercultural sensitivity to develop. However, these choices cannot be based on either absolute or universal principles. Rather, *ethical behavior must be chosen with awareness that different viable actions are possible.* (Bennett, 1993:6) (italics added) (p. 281).

The level of reflection and ethical development that the process requires is extremely high. It is not inconceivable that the high school students involved in this study, some as young as grade nine, might have achieved the requisite levels of moral and cognitive development to consciously construe difference, to self-consciously reflect on their experiences, and to make ethical choices, but there was no evidence in the article to suggest that these considerations entered into the choice of the IDI as an appropriate instrument for this study, or the DMIS as an appropriate theoretical framework.

The issue of consequential validity is important here in two ways. One way to consider it is in terms of the IDI's authors' assurances that their instrument is reliable and valid. These assurances, from two very respected figures in the field of cross-cultural research, are what prompted me (and Westrick, I suspect, since a researcher can use the IDI only after attending a qualifying workshop given by the authors) to reach for the IDI as the empirical solution to my research problem.
The second way to look at consequential validity is in terms of the user of the instrument. Although the decision to use the IDI may have originated in the misconception that the 'instrument is valid', Hambleton and Patsula (1998) remind us that,

(A) researcher using an adapted test still has the responsibility of producing evidence of validity in the context where that adapted test is used" because "researchers risk imposing conclusions based on concepts which exist in their own cultures but which are foreign, or at least partially incorrect, when used in another culture (p. 156).

Hambleton and Patsula refer to an adapted test, but in the case at hand, while it was the subjects who were 'adapted' (i.e., different from the population upon which the original data inferences were validated), the same principle applies.

I am not trying to imply that Westrick's 2004 study is necessarily flawed, but it is deeply disturbing that validity issues were not even considered, neither in the traditional, practical sense (as a property of the inferences drawn from the data), nor in terms of consequential validity; the appropriateness of using the framework and the instrument in the particular context. Most practitioners do not question validity claims, and would be stupefied to discover they needed to address validity issues every time they wanted to use an instrument in a novel context. Yet this is precisely
what Hambleton and Patsula, and other experts in validity (see Messick, 1995, 1998; Hubley & Zumbo, 1996), are calling for.

The only validation studies to date (Bennett & Hammer, 1998; Paige et al., 2003) have used data and inferences from respondents in the United States, using the original English version. This would not be a shortcoming in and of itself, were it not for the issue of consequential validity and practitioners' and researchers' propensity to believe that an instrument can be validated if rigorous statistical protocols are followed.

Messick (1995) laid out the validity issues that shaped this research as follows:

Validity is not a property of a test or assessment as such, but rather of the meaning of the test scores. These scores are a function not only of the items or stimulus conditions, but also of the persons responding as well as the context of the assessment. In particular, what needs to be valid is the meaning or interpretation of the scores; as well as any implications for action that this meaning entails (Cronbach, 1971). The extent to which scores' meaning and action implications hold across persons or population groups and across settings or contexts is a persistent and perennial empirical question. This is the main reason that validity is an evolving property and validation a continuing process (p. 741).

The results of this study confirmed that validity is in part dependent on the persons responding to the instrument, that conclusions based on concepts from one culture may be at least partially incorrect when used in another culture, and
that validation is an ongoing process. The inferences that one must draw from analyses of data from the Japanese-language version of the IDI administered to Japanese subjects differ significantly from those drawn with the original IDI and English-speaking subjects in the United States. The discussion surrounding information from native-Japanese-speaking informants, in the translation process and elsewhere, will highlight that.

The ongoing process of validating the IDI in translation was also significant because although the IDI is positioned as an instrument for measuring intercultural sensitivity it, as an extension of the DMIS, can more broadly assess orientations towards difference of all kinds. However one defines the culture with which he or she identifies, including sexual orientation, linguistic group, ethnic group, or age group, to name a few, the IDI potentially offers a window into a person’s worldview vis-a-vis those different from him or herself. Thomas and Inkson (2004) described culture in their book on Cultural Intelligence in the broadest terms. "'Culture' in this case is not confined to national or ethnic culture, but, consistent with our definition in this book, can be any social group, and subcultures in a society . . ." (p. 73). The IDI’s potential value in helping us to understand orientations toward and (in)tolerance of difference make it
worthwhile to rigorously examine its cross-cultural transferability and to work towards modifying it, as necessary, to make it work seamlessly across cultures.

The DMIS is not all that helpful in explaining how one's worldview shifts from one stage to the next, except to say that construing and reconstruing one's experience through critical reflection trigger change. This does not, however, diminish its value in providing a framework for describing the progression from an ethnocentric to ethnorelative worldview. The work of researchers such as Matsumoto et al. (2003), Chen (1997) and Chen and Starosta (2000) may help us to understand the substructure or constituent components of intercultural understanding.

4.2 The Validation Process

According to Messick (1995) validity has traditionally been seen to consist of three components: content, criterion-related, and construct. Angoff (1988, cited in Hubley & Zumbo, 1996), reported that although currently accepted thinking on validity is that content, criterion-related, and construct validity are "three aspects of a unitary psychometric divinity" (p. 211) Hubley and Zumbo (1996) commented on common practice in the field as follows:

An examination of studies reporting on the validity measure or observation indicates the presence of essentially three assumptions related to the
traditional view of validity: (a) validity is a property of the observation or measure (italics in original), (b) there are various types of validity, and (c) reliability and validity are presented as distinct concepts with different purposes (p. 209).

The validation work conducted by Bennett and Hammer (1998) and later by Paige et al. (2003) were products of the three assumptions cited above, with particular emphasis on the first two; that validity is a property of the instrument and that there are various types of validity to be tested.

The following discussion, while not accepting either the property-of-the-instrument, nor the discrete-elements assumptions about validity, was set out within the framework of the original validation work.

4.3 Content Validity

Content validity in the original IDI was established following the guidelines for test construction set out in DeVellis (1991, as cited in Hammer, Bennett, and Wiseman, 2003). It was, to all appearances, a thorough and rigorous process. While I am not criticising the procedures followed, that very rigour, set in a context of binary decisions regarding validity (i.e., that an instrument is or is not valid) might have encouraged certain conclusions to be drawn that appear to have been unwarranted.

Ignoring the validity issues inherent in language and culture adaptation, the IDI's authors clearly intended the
instrument to be interculturally transferable. They (Hammer, Bennett, & Wiseman, 2003):

were initially concerned that the empirical observations upon which the DMIS was based could be re-created in systematic ways. This concern was addressed by examining discourse of people from a variety of cultures in order to determine if observers could reliably categorize the discourse in ways identified in the DMIS theoretical framework (p. 7).

Care was taken to have a broad cultural and experiential base. Hammer et al. (2003) report that:

While the pilot interviews were conducted with individual students from a variety of cultures, it was decided that the actual sample of interviewees would consist of people of varied cultural backgrounds who also extended beyond the university community. Therefore, the interview sample was selected from residents from such places as the International House in Washington D.C. (where professionals from many different countries reside) as well as various places of employment in and around the Washington D.C. area (p. 7).

Forty people as described above were interviewed regarding their experience with, and attitudes toward, other cultures. Of the forty, there were twelve U.S. Americans of European origin, three more U.S. Americans of South-Asian origin (bringing the number of U.S. Americans to fifteen of the forty). There were also three people from each of Britain, Japan and France, two from each of Switzerland, Korea, Ireland, and Russia, and one person from each of China,
Denmark, Spain, France, Germany, Estonia, India, Turkey, Ecuador, Guyana, and Ivory Coast.

There were a number of ways to interpret these data besides the conclusion drawn by Hammer et al., that they had a sufficiently culturally-diverse mix to assure the cross-cultural robustness of the IDI. Most importantly in my view, all forty subjects spoke English sufficiently well to participate in a lengthy interview of some conceptual sophistication; a discussion of their experience of cultural difference. Had the non-native speakers been interviewed in their native languages and quotes from the translations of those transcripts been used in the analysis, a different set of items might well have emerged, if not a DMIS of a different texture.

This conclusion was not merely speculative. Yamamoto (1998) used the IDI interview protocol to interview Japanese university students studying in the United States, in Japanese. First, she analysed the interview data against the DMIS stages. Then, she classified the data into the seven categories that she found emerging naturally from them: “Attention to Physical Difference; Physical Admiration for Caucasians; Attention to Physical Similarity; Attention to Own Frame of Reference; Naturalness of Difference; Inevitability of Difference; and Suspension of Judgement” (p. 77). She
states that "(T)hese emergent categories are closely related to Japanese cultural values and perceptions of reality" (p. 77). Going on to compare the DMIS to her emergent categories, she found that IDI:

statements such as 'I appreciate (enjoy or respect) cultural differences' were hardly expressed by the Japanese students. . . Also, the students paid much attention to the differences and similarities in physical appearance, which they associated with the degree of discomfort or comfort (p. 77).

Yamamoto concluded that:

These results suggest that the definitions of each stage may need some modification in order to understand intercultural sensitivity in the Japanese context. It might be possible to say that what Japanese perceive as differences/similarities or how they deal with difference/similarities are different from or not included in the stages of the model. These aspects need to be considered and added to the model in order to modify it to apply in the Japanese context (pp. 77-78).

I was also reminded here that when I was going over my RA's responses to both the English and Japanese versions of the IDI, her responses differed in places where items that were problematic in Japanese were conceptually available to her in English. Additionally, one of my translators, commenting on a particular item and concepts differentially available in English and Japanese said, "Another mutsukashii (difficult) sentence, referring to culture in terms of superior and inferior, although it makes sense in English" (personal communication).
Returning to the subject pool, another way to look at it is to note that its members are overwhelmingly from Judeo-Christian backgrounds, assuming that the twenty-eight subjects of European descent were members of the mainstream. Socialisation into a particular religio-cultural worldview has obvious implications in the context of tolerating, appreciating, or enjoying difference.

For the next step in the instrument-building process, culling utterances to include in the IDI, four members of the research team each reviewed twenty-five randomly selected transcripts from among the original forty. Hammer et al. (2003) reported that they "rated the DMIS orientations the interviewees' (sic) most consistently expressed during the interview" (p. 8). These resulted in an item pool of 239 IDI sample items for further pilot testing.

Although adequate inter-rater reliabilities were obtained in this process, ranging from .66 (fair) to .86 (excellent), we do not know how many utterances survived from non-U.S. Americans, for example, or non-native speakers of English. It could very well be that the utterances that consistently reflected the DMIS orientations during the interview came from particular cultural or linguistic subsets of subjects; a threat to validity that Messick (1995) calls "construct underrepresentation" (p. 742).
Another potential shortcoming of the content-validation process was that although Hammer et al. (2003) tell us that the pool of 239 quotes resulting from the process outlined above was twice tested in a pilot version of the IDI, "with a culturally diverse group of people" (p. 9) the pilot testing was looking for "clarity of instructions, item clarity, response option applicability, and overall amount of time taken to complete the instrument" (p. 9), but not conceptual transferability.

The surviving items (assuming that some of the items failed the test of item clarity and response-option applicability) were then given to another group of DMIS experts who further distilled them into the sixty items that comprise version one of the IDI. This methodology was not without redeeming features. It had the twin merits of employing items that were not generated by experts sitting down to write what seemed to be reasonable items for field testing, and a statistically-rigorous process involving agreement by experts on the strength of the relationship of the items to the DMIS. However, it lacked rigour in its failure to confirm the cross-cultural robustness of the items.

While the authors claim that the use of a culturally-mixed pool of interviewees demonstrates pan-cultural content validity, in addition to simple content validity, I believe
that the arguments raised in this discussion were sufficient to bring that into question. Further research with non-Indo-European language versions and with subjects from other than Judeo-Christian backgrounds is certainly warranted before the pan-cultural applicability of the IDI can be asserted with any confidence.

4.4 Criterion-related Validity

In the original validation work, criterion-related validity was tested by comparing the results of the IDI with scores obtained on two other scales, the Worldmindedness Scale and the Intercultural Anxiety Scale. A third scale was used to assess social desirability effects (Bennett & Hammer, 1998). Bennett and Hammer (1998) report that "(T)he Worldmindedness Scale (Sampson & Smith, 1957) is a measure used to assess "internationalistic" attitudes" (p. 59). A six-item version of the scale was used for the original validation. Results obtained (see Table 4.1, reproduced from Bennett & Hammer, 1998, p. 81) showed that higher levels of Worldmindedness correlated negatively with ethnocentrism and positively with ethnorelativism. Half of the correlations (.40 and above) are considered strong in a strictly statistical sense (although they account only for between sixteen and twenty-four percent of the variance) the other half are quite weak and there is no
information provided on how well the Worldmindedness Scale 'travels' interculturally. This might not have been significant for the original IDI with English-speaking subjects. However, any claim of cross-cultural transferability cannot be legitimately made without empirical evidence that there is no language or culture effect on the instrument's performance. Since the criterion-referenced evidence is correlational that must include instruments that are introduced to establish criterion-related validity through correlations.

A modified version of Stephan and Stephan's (1985) Social Anxiety Scale, focusing "on the degree of anxiety respondents experience when interacting with people from cultures other than their own" (Bennett & Hammer, 1998, p. 81) was used as a second measure. Bennett and Hammer (1998) report that:

A number of studies have found this measure to maintain satisfactory reliabilities across cultural contexts (e.g., Gao & Gudykunst, 1990; Gudykunst, 1989; Hammer, Wiseman, Rasmussen, & Bruschke, 1998) (p. 81).

However, as Table 4.1 shows, the correlations between the Intercultural Anxiety Scale and the IDI are not particularly strong, even in a purely statistical sense, and the relationship between the Minimization Scale and the Intercultural Anxiety Scale is obscure.
Table 4.1 Correlations between IDI and Worldmindedness and Intercultural Anxiety Scales

<table>
<thead>
<tr>
<th>IDI Scale</th>
<th>Worldmindedness Scale</th>
<th>Intercultural Anxiety Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denial</td>
<td>-.40, p=.001</td>
<td>.28; p=.001</td>
</tr>
<tr>
<td>Defense</td>
<td>-.18; p=.007</td>
<td>.13; p=.05</td>
</tr>
<tr>
<td>Minimization</td>
<td>-.18; p=.007</td>
<td>-.11; p=.08</td>
</tr>
<tr>
<td>Acceptance</td>
<td>.47; p=.001</td>
<td>-.17; p=.01</td>
</tr>
<tr>
<td>Cognitive Adaptation</td>
<td>.47; p=.001</td>
<td>-.21; p=.002</td>
</tr>
<tr>
<td>Behavioral Adaptation</td>
<td>.36; p=.001</td>
<td>-.15; p=.02</td>
</tr>
</tbody>
</table>

(reproduced from Bennett & Hammer, 1998, p. 81)

Bennett and Hammer (1998) concluded that the results obtained from the Worldmindedness and Intercultural Anxiety scales "offer compelling evidence of the construct validity of the IDI" (p. 82). I would beg to differ. As noted above, the criterion-related validity evidence produced for the original IDI was not convincing, except to confirm a general pattern.

Because I didn't have the resources to duplicate the process of translating the other two scales as I had with the IDI, I was originally prepared (when I embarked on this project) to accept the criterion-related validity evidence from Bennett and Hammer's (1998) original work as sufficient to make an argument for the criterion-related component of the validity evidence. It is the only evidence available and it certainly seemed to satisfy the authors. Fortunately, the work
I've done in examining conceptions of validation in this study has precluded me from taking the easy way out. The wishy-washiness of the results is compounded, for my purposes, by the lack of evidence that the Worldmindedness Scale has any cross-cultural applicability. One could be excused for saying that it's a seemingly impossible task to obtain convincing criterion-related validity evidence involving a number of instruments that have been shown to yield consistent results across cultures, but that is exactly the level of evidence that must be adduced before claims of cross-cultural transferability can be made.

4.5 Social Desirability

Bennett and Hammer's (1998) report on the validation process mentioned the addition of a scale to test whether social desirability was a factor in IDI responses, but they did not report the results for social desirability. In IDI workshops and other fora, however, Bennett and Hammer state that social desirability is indeed not a factor. They did present statistical evidence against any effects from gender and socio-economic status, as measured by level of education, in their validation results (Bennett & Hammer, 1998). The influence of gender in Japanese social roles has been widely noted. Hofstede (1980) described Japanese culture as being the
most 'masculine' in his study; the culture in which gender roles and behaviour were most clearly and rigidly differentiated. As a consequence, I could not assume that the results of analysis for gender effects in the original validation study would necessarily hold for a Japanese sample so I undertook an analysis from my own data, based on mean IDI scores. There were an insufficient number of cases (particularly with an n of 106 for male respondents) to redo a principal components analysis on the basis of gender. The analysis undertaken showed no gender effect (see Table 4.2).

Heine and Lehman (1995) have done some work on the social desirability aspect of the validation procedure. They compared Canadian and Japanese students' responses on the Balanced Inventory of Desirable Responding (Paulhus, 1991), using a Japanese translation for the Japanese subjects, and concluded, tentatively, that:

In the context of an anonymous questionnaire, then, there was no evidence that the responses of the Japanese were more socially desirable than those of the Canadians. A tentative conclusion is that comparisons of the responses of Japanese and North Americans—at least for students—on anonymous questionnaires are not confounded by socially desirable response sets (p. 779).
Table 4.2 Analysis for Gender Effect by Mean Score

<table>
<thead>
<tr>
<th>Between-Subjects Factors</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENDER</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>106</td>
</tr>
<tr>
<td>F</td>
<td>287</td>
</tr>
</tbody>
</table>

Tests of Between-Subjects Effects

Dependent Variable: MEAN

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>7.181E-03</td>
<td>1</td>
<td>7.181E-03</td>
<td>.010</td>
<td>.919</td>
</tr>
<tr>
<td>Intercept</td>
<td>8781.229</td>
<td>1</td>
<td>8781.229</td>
<td>12602.069</td>
<td>.000</td>
</tr>
<tr>
<td>GENDER</td>
<td>7.181E-03</td>
<td>1</td>
<td>7.181E-03</td>
<td>.010</td>
<td>.919</td>
</tr>
<tr>
<td>Error</td>
<td>272.452</td>
<td>391</td>
<td>.697</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>11408.500</td>
<td>393</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>272.459</td>
<td>392</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*R Squared = .000 (Adjusted R Squared = -.003)

This seemed to minimise concerns about possible confounding effects from social desirability responses, but Heine and Lehman (1995) also cited studies that concluded that:

Cultural differences have been observed with respect to a moderacy response bias; answers given by Japanese respondents to a questionnaire tended to be closer than those of U.S. respondents to the midpoint on Likert-type scales (Stening & Everett, 1984; Zax & Takahashi, 1967) (p. 777).
This was significant for two reasons. The first was that it again reinforced the need to empirically verify the applicability of the criterion-related validity measures used in the original process (both of which employed Likert scales). The second, and more important reason, was that the midpoint responses to the IDI are critical to interpreting the data.

After reverse coding of negatively worded items, each item in the IDI has an ideal score in terms of the perfectly interculturally-sensitive person of one, strongly disagree, for ethnocentric items and seven, strongly agree, for ethnorelative items. In analysing IDI data, scores of one or two, six or seven are non-problematic as they represent clearly-held convictions. Scores in what Hammer calls 'the bucket', in the three to five range, identify where the respondent is 'having issues'. Responses indicating uncertainty about an item are key indicators about where the respondent is working through ambivalence in the corresponding DMIS stage.

If Japanese respondents are actually more likely, as a cultural artefact, to choose midpoint items on Likert-scaled instruments, that cultural tendency could compromise the interpretations made from the data and represent a clear
threat to validity. This is another area in which further research is clearly called for.

4.6 Construct Validity

The principal components analysis yielded important insights both into the IDI and also into the validation process itself. The number of subjects in this study was larger than the group used by Paige et al. (2003) in their analysis (n=330) and also larger than the original IDI validation study (n=226). It was noteworthy that seventeen of the sixty items, 28.33% of the instrument, did not map onto their predicted IDI stages. That indicated a need for a deeper analysis of those items, but of a culturally and qualitatively different kind than that which comprised the validation process for the original or the replication by Paige et al. (2003).

In addition to items that did not map onto the expected factor, there were a number of items that loaded with some strength on a number of factors across the matrix, indicating that their relationship with the underlying DMIS stages was ambiguous, further clouding the issue of the Japanese IDI's construct validity. Among the items that loaded above the threshold (minimum of .30) on more than one factor, were (see Table 3.3 - Rotated Factor Matrix) item 44 which loaded at .377 on Factor 4 (Denial) in addition to its main (or
strongest) loading of .514 on the Defence factor. Item 27 loaded at .343 on Factor 1 (Adaptation) in addition to its main loading of .470 on the Acceptance factor. Item 34, an orphaned Minimization item in the middle of the Acceptance factor (with a main loading there of -.460) had a secondary loading of nearly equal strength of .419 on the Minimization factor. There were also a number of items that loaded on other factors with some strength, although the loadings were below the .30 threshold. Item 37, a Denial item marooned in the Adaptation factor loaded secondarily on Denial, but only at .286. Item 58, a weak Adaptation item (loading of .331) loaded almost as strongly (.254) on Factor 3 (Acceptance). Item 30 loaded (somewhat weakly) on its 'home' Denial factor at .364, but had two other loadings of nearly the same strength (.298) on Factor 2 (Defence) and on Factor 3 (.266), the Acceptance factor. These ambiguous items bring the total to 40% of the IDI items in the Japanese translation that do not perform as expected vis-à-vis results from the analyses on the English original.

While the fact that some of the items did not perfectly map onto their intended IDI stages was not a novel finding, it was interesting to follow how the lack of fit has been variously interpreted.
Consider the Minimization stage of the IDI as example. In the Japanese translation, the three Transcendent Universalism items clustered as a factor unto themselves (see Results). They had been flagged during the translation process as being potentially problematic on a conceptual level and that prediction was borne out in the principal components analysis. We can trace similar results through the original validation process (Bennett & Hammer, 1998). Bennett and Hammer's (1998) original factor analysis:

resulted in a ten-item scale with reasonably high corrected item-total correlation and a reliability estimate of .87 (n=212). These items reflect a mixture of both "Physical Universalism" and "Transcendent Universalism" within the Minimization stage description of the DMIS and therefore were labelled with MINIMIZATION (p. 67).

So, in the original analysis of the factor structure, the items in Minimization were deemed to be sufficiently correlated and coherent to form a single dimension with two subscales.

When Paige et al. (2003) performed their analysis, their conclusion was also that while:

... the analyses of the internal structure of the IDI have shown it to be a reasonable approximation of the theoretical model of intercultural development. Minimization items split along the theoretical lines of physical and transcendental universalism (Factors 3 and 4) (p. 483).
They went on to say that, "(T)he Minimization split is interesting because it follows exactly the form structure of the DMIS" (p. 484) and speculated that their subjects had responded to the Transcendence items the way they had because "it is very likely that many individuals of the age groups represented in our sample have not given any serious thought to their position on the issues of spirituality and religious beliefs" (p. 484).

In reporting on the Paige study, however, Hammer, Bennett, and Wiseman (2003) interpreted the findings to mean that:

the factors identified in the 60-item IDI might not be as stable as desired. Further, their findings suggested the possibility of three more fundamental dimensions: a factor composed of Denial and Defense items, a Minimization factor, and a factor that largely consists of Acceptance and Adaptation (pp. 11-12).

This is somewhat surprising given that Paige et al. (2003) reported that, "(O)verall, the factor analyses provide strong empirical support for the broader two-factor (ethnocentric and ethnorelative) structure of the development model" (p. 483), but all results are open to interpretation.

On the surface, the results from the principal components analysis of the Japanese IDI and those obtained by Bennett and Hammer (1998) and Paige et al. (2003) were in agreement. All three showed a split structure for Minimization with
Transcendent Universalism clustering separately from Physical Universalism.

However, my Japanese informants (the translators, my RA, and the raters who helped me to refine the items), unanimously noted that the Transcendent Universalism items were conceptually nonsensical in Japanese culture, despite the fact that the words could be translated. As one of the translators put it in private e-mail correspondence "I don't understand this whole sentence . . . M [another translator] avoided translating all 'the children of a spiritual being' [items]. I don't know what that is either. It's possible to put Japanese words together as a translation without understanding what that means. demo hen" [but it's weird — author's translation].

Yamamoto and Tanno (2002) cited a similar difficulty. They said, (in translation from the original Japanese by this author) that:

I made an effort to be as faithful to the original wording of the IDI as possible, but some expressions came out as difficult to understand in translation. However, it wasn't a simple translation problem. It is possible that some of the items were difficult to understand culturally. For example, in the Universal Values subscale of the Minimization stage items 12, 23 and 59 are phrased as 'At our root is a supernatural holy being (choshizende shinseina sonzai), but the English original for 'supernatural holy being' was 'spiritual being'. I thought about translating it [in a syllabic katakana rendition] as 'supirichuaru bi'ingu', but trying to be faithful to the original, I checked with Bennett and as a result I used 'supernatural holy being'. However, some doubt remains as to whether the value held by those who
believe we are united under a supernatural being is an appropriate expression of what the Minimization stage of intercultural sensitivity means to Japanese (p. 40).

Although the most glaring examples were the items that comprised the Transcendent Universalism subscale of Minimization, clearly products of a Judeo-Christian mainstream culture, this was not a quibble over three items out of sixty. It was fundamental to the way the instrument was constructed; out of quotes from actual interviews, not fabricated items.

Bennett and Hammer (1998) and Paige et al. (2003) looked at the data through their own cultural and linguistic lenses and concluded that they were an accurate reflection of the DMIS. If that conclusion had been limited to the English version of the IDI, administered to highly-proficient speakers of English, there would be no problem. However, I believe Bennett and Hammer were being premature at best in pronouncing the IDI to be valid, in any language version and with any cultural group, because they had used a culturally and linguistically diverse pool of subjects in the item-building phase. From the results I obtained and the conclusions drawn by Yamamoto (1998) it is clear that had the original interviews been conducted in Japanese, for example, the items appearing on the IDI would have been at least somewhat, perhaps substantially, different.
Other issues came to light during consultations on the translation. One recurring issue was the IDI’s references to 'being a member' of one’s own culture as juxtaposed with people from other cultures. Although some of the phrases sound (to my ear) a bit laboured even in English, e.g., ‘Although I feel I am a member of my own culture . . .’ they are otherwise unremarkable in English. To the Japanese, references to one’s own and other people’s cultures have an unnatural clang. As one translator put it, “I don’t think Japanese people distinguish people by culture, but by country or nationality” (personal correspondence). In the Japanese worldview there are two types of people in the world, nihonjin (Japanese) and gaijin or gaikokujin (literally ‘outside’ people or ‘foreign-country’ people), i.e., non-Japanese. This is not dissimilar, in my personal experience, to the Jewish view of the world as consisting of Jews and goyim (non-Jews). Finer cultural distinctions than that are generally not germane. It is difficult to say what effect the specific insistence on the word ‘culture’ (to which I remained faithful in the Japanese version of the IDI) might have had on respondents. One translator had to consciously make an effort not to substitute gaikokujin for the phrase ‘people from other cultures’ and was at times overcome by the unnaturalness of the construction to her Japanese sensibilities. It certainly made the items stand
out, compelling the Japanese respondents to either reflect on the relationship in their minds between nationality and culture or to automatically substitute the more natural gaikokujin in their minds for 'people from other cultures'. Another unintended consequence might have been to merely annoy respondents and weaken the 'face validity' of the instrument.

It might be this proclivity for dichotomising the world that led three of the Denial – Disinterest subscale items and one Minimization – Superficial Differences item to fall into factor four, which was primarily an Acceptance (seven out of twelve items) factor, in my analysis. There are four Acceptance subscales; Value Relativity, and Enjoying Difference, Describing Difference, and Learning Difference. It seems that for Japanese respondents, there might have been a 'Difference' factor rather than an Acceptance factor. This notion was reinforced by Yamamoto's (1998) list of dimensions that spontaneously emerged from her interviews in Japanese with Japanese subjects. In her analysis, three of the seven dimensions, Attention to Physical Difference, Naturalness of Difference, and Inevitability of Difference cited difference specifically, with another, Attention to Physical Similarity, invoking difference as its opposite.

This was particularly contentious for construct validity with this population because Denial and Minimization lie on
the opposite side of the ethnocentric-ethnorelative divide
from Acceptance and so, unlike items from adjacent dimensions
crossing factors, are theoretically incompatible within the
same factor.

Had I relied entirely on formulaic validation protocols
consisting of inter-rater reliability statistics and the
numerical rules of thumb for factor analyses I might have
drawn the same conclusions as previous validation studies;
particularly if I had been working from the premise that the
instrument, rather than the inferences made with it, could be
validated. However, interpreting the results of the principal
components analysis through the lens of information from the
translation processes made that impossible.

4.7 The Translation Process

My understanding of the data from the Japanese IDI would
have been much poorer had it not been for the translation
protocols employed.

The contribution this study makes to the field of
translating questionnaires and instruments is to suggest two
additional steps.

The first was to use the translation outside of the
translation framework. The most thoughtful approaches to
translation (and adaptation) involve intensive discussions
over the accuracy, naturalness, and acceptability of the translations. These, however, take place within the translation process itself, limiting the focus and scope of the discussion. The step that I added was outside of the translation cocoon; using the draft translation as a rating tool for 'amateur' attempts to translate the items (step three of Table 2.1, above). The distinction might be subtle, but analysing translation options, word choice, syntax, etc., is a self-referential exercise. It exists within the context of the translation. Taking a step outside of the translation process to see how the items worked in the context of interpretations by the instrument's intended final users led to different insights in terms of how non-translators might interpret wording or, more importantly, concepts.

The second contribution this study made was to introduce the step (step five in Table 2.1) of having a bilingual complete both versions of the instrument consecutively. Ideally, if the instrument performs equally well in both cultures and the translation is 'perfect' the score for each item would be identical. Any discrepant scores would point to a problem either with the instrument or the translation. This procedure uncovered some minor errors in the translation. More importantly, however, score discrepancies confirmed the existence of conceptual issues, as noted above.
5.1 Contributions to the Field

This research has contributed to the field in a number of ways.

The validity issues raised in this paper should have real resonance for practitioners, and even for academics in the field who reach for an instrument because they believe 'it is valid'.

It was the first study to examine the validity of inferences made with the IDI in translation, with a culturally-different population. In doing so it raised strong doubts about the cross-cultural transferability of version one of the IDI and raised some questions about the DMIS as a model for understanding worldviews with respect to difference, in cultures other than U.S. American, and perhaps even within U.S American culture itself, given its breadth and diversity. By introducing a culturally-based alternative analysis to what appeared to be similar data sets, the study highlighted the importance of current thinking in validity theory that instruments cannot be validated, only specific inferences made from data. Issues related to consequential validity in the use of an instrument or a theoretical framework in contexts other...
than the ones for which they were specifically developed were also raised.

The study has also contributed to the field of translation by introducing additional steps to enrich translation protocols.

5.2 Limitations of the Study

The study was limited primarily by its subject pool. All of the subjects were Japanese university undergraduates, within a fairly narrow range of age and cross-cultural experience. Given the role of critical self-reflection in advancing intercultural sensitivity, subjects with a broader life experience might have responded differently to the IDI items.

The study was also limited in a deliberate way by its self-imposed literal faithfulness to the IDI. This seemed, and seems, necessary in terms of testing the claimed cross-cultural transferability of the instrument, but further research with a truly adapted, as opposed to a translated version of the IDI, would speak more clearly to its actual potential for generalisability.
5.3 Directions for Further Research

A number of research projects are suggested by this study. The first would be to replicate the study with an adapted, rather than a translated version of the IDI, as suggested above. The adapted version could profit from an expanded subject pool, encompassing a greater range of age and experience, both life experience and cross-cultural experience. Research of this nature on an adaptation of version two of the IDI would be of even greater utility.

Further research in the direction suggested by Yamamoto (1998), looking at what Japanese respondents say in relation to their experience of cultural difference and comparing that to the DMIS, is also warranted. If Japanese speakers did not reproduce the range of stages suggested by the DMIS, or clearly produced a stage or stages not presently encompassed by the model, the DMIS would have to be rethought, at least in terms of its potential to more generally illuminate the human condition.

Another, related, area to pursue is a more systematic examination of whether a 'Difference' dimension actually exists for Japanese respondents and the implications that a dimension encompassing both ethnocentric and ethnorelative elements might hold.
5.4 Implications for Practice

Given that this dissertation was written from a practitioner's perspective, I hope that the implications for practice have been salient throughout. To make them explicit, this study will serve as a caution to other practitioners who might be tempted to reach for a psychometric or other instrument in the belief that it has been validated and is therefore suitable to the purpose they have in mind. Practitioners need to be cognizant of their responsibility to ensure that whatever instrument they choose, or research design they formulate, satisfies the criteria for construct validity or other standards of appropriateness for their particular research contexts. This is particularly true for practice because those results are likely to be translated into programmatic action.

Those working across languages (in the case of translated or adapted instruments) or cultures (even if using a research instrument in its original form) face a particularly onerous burden of proof since they are, in effect, navigating uncharted waters.
5.5 Afterward

This research sprang from an attempt to define and quantify a tangible facet of internationalisation for students to take home with them from an exchange experience. Although intercultural sensitivity remains a worthy goal we appear to be only slightly closer to operationalising and quantifying it. In the process of conducting this research, I have learned a lot about the notion of validity, in cross-linguistic and cross-cultural contexts in particular, and I have come to understand the DMIS in a more profound way. I hope and believe that both of these deeper understandings will return to enrich my practice, bringing the exercise full circle.
BIBLIOGRAPHY


Yamamoto, S., & Tanno, D. (2002). *Ibunkakanjuseihattatsushakudo (The Intercultural Development Inventory) no nihonjin ni taisuru tekiyohsei no kentoh: Nihongo bansakusei wo shiya ni irete (Assessing the applicability of the Intercultural Development Inventory to Japanese: bringing in the viewpoint from the creation of a Japanese version).* Journal of the Aomori National University, 7 (2), 24-42.
