

**Clarifying Inaccurate Statements Characterizing the Intercultural Development
Inventory in Matsumoto and Hwang (2013) JCCP Article¹**

by

Mitchell R. Hammer, Ph.D.

Professor Emeritus, American University, Washington

Principal, IDI, LLC

August 20, 2015

Contact information:
Mitchell R. Hammer, Ph.D.
P.O. Box 1388
Berlin, MD 21811
mhammer@idiinventory.com
phone: 443-664-5350

¹ The Author acknowledges Peter J. Jankowski, Ph.D. Associate Professor, Bethel University, St. Paul, Minnesota for insights incorporated in this article.

Abstract

In this reply, I present specific concerns regarding the following article published in the *Journal of Cross-cultural Psychology*: Matsumoto, D. and Hwang, H.C. (2013) "Assessing cross-cultural competence: A review of available tests," *Journal of Cross-cultural Psychology*, 44, (6), 849-873.

I first offer a response to Matsumoto and Hwang's criteria for evaluating cross-cultural competence, specifically in terms of (1) lack of reliance upon and citation of established criteria for evaluating measures, (2) the inappropriate grouping of all ten instruments under one global statement of intended use and purpose, and (3) oversight of a tautological problem, especially regarding ecological validity. This reply concludes with specific clarifications of inaccurate statements made by Matsumoto and Hwang regarding the validation of the Intercultural Development Inventory (IDI).

Clarifying Inaccurate Statements Characterizing the Intercultural Development Inventory in Matsumoto and Hwang (2013) JCCP Article

Preface

The Journal of Cross-cultural Psychology (JCCP) published an article (Matsumoto and Hwang, 2013) that purported to offer a transparent and impartial review and evaluation of measures of cross-cultural competence. The Intercultural Development Inventory (IDI) was one of ten instruments assessed by Matsumoto and Hwang. After reading the article, I was dismayed to find a very substantial conflict of interest of the authors which was not acknowledged, per JCCP policy. JCCP did subsequently print the following conflict of interest clarification: *“David Matsumoto is the developer of the Intercultural Adjustment Potential Scale (the “ICAPS”), one of the tests reviewed in the article. Dr. Matsumoto is additionally founder and director of Humintell, LLC, which sells the ICAPS test. Hyisung C. Hwang is a Research Scientist at Humintell, LLC.*

This clarification is not sufficient in specificity. The conflict of interest Matsumoto and Hwang have is not simply a “potential” conflict of interest because Matsumoto developed and markets ICAPS through his for-profit company; it is that he is on record as concluding years before his 2013 “evaluation of instruments” article was published in JCCP that his own instrument is the only test that validly measures what he describes as cross-cultural competence in the 2013 article. The following statement appears (documentation of this fact is available) on their company’s (Humintell) website (www.humintell.com) early in 2010 and continues each subsequent year: *“the ICAPS is the only test available that reliably and validly measures the psychological skills associated with successful intercultural and intracultural adjustment.”* In another document at: www.davidmatsumoto.com, the following is stated as early as 2009: *“The ICAPS-55 is the only scale available today that can reliably and validly predict the degree to which a person will successfully adjust to living, working, and playing effectively in a new and different cultural environment.”* What is not acknowledged in the clarification of conflict of interest statement recently issued by JCCP is that this conclusion of the superiority of their own ICAPS instrument was already made on the Humintell website well before the Article was submitted for publication in JCCP. These de facto conclusions made years before their evaluation of instruments in their 2013 article clearly challenge the impartiality of their assessment in their 2013 publication.

A second concern the reader should be made aware is that I requested JCCP to publish this reply (without the remarks in this preface). The Editor refused to send my reply article out for peer-review. Therefore, I offer an alternative interpretation and clarifications regarding the treatment of the IDI by Matsumoto and Hwang (2013) in this document.

Introduction

In this reply, I present my concerns regarding the following article published in the *Journal of Cross-cultural Psychology*: Matsumoto, D. and Hwang, H.C. (2013) "Assessing cross-cultural competence: A review of available tests," *Journal of Cross-cultural Psychology*, 44, (6), 849-873. This reply article draws upon but is distinct from Jankowski's (under review) comprehensive assessment of validation testing of the IDI following AERA/APA/NCME (2014) test construction criteria.

I offer a response to Matsumoto and Hwang's criteria for evaluating cross-cultural competence, specifically in terms of (1) lack of reliance upon and citation of established criteria for evaluating measures, (2) the inappropriate grouping of all ten instruments under one global statement of intended use and purpose, and (3) oversight of a tautological problem, especially regarding ecological validity. I then conclude with clarifications of inaccurate statements made by Matsumoto and Hwang regarding the validation of the Intercultural Development Inventory (IDI).

I developed and make available the IDI to educators, trainers and researchers from throughout the world. In this reply, therefore, I focus my attention on Matsumoto and Hwang's treatment of the validation undertaken on the IDI vis-à-vis their evaluation criteria, statements made regarding the IDI, and subsequent conclusions drawn from their evaluation.

Concerns Regarding the Criteria Used to Evaluate 3C Measures

I offer three observations of the article by Matsumoto and Hwang (2013) and their interpretation of the evidence available for evaluating the IDI specifically. I acknowledge that there is value in compiling published empirical evidence on measures of cross-cultural competence (3C). A presentation of the available evidence in one place can be useful to researchers trying to decide which measure best fits their research question. I also acknowledge that their compiling of the evidence was also *evaluative*, and hence they offered an interpretation "that the CQ, ICAPS, and MPQ have the most promising evidence for assessing 3C" (p. 867).

First, I respect the authors' right to offer an interpretation of the empirical evidence available on the different measures of 3C, and at the same time acknowledge, that it is just that, an *interpretation*. The evaluation of a measure clearly involves "professional judgment" (AERA/APA/NCME, 2014, p. 1). As such, this implies that multiple interpretations of the evidence for or against a particular measure could be constructed. The difficulty in evaluating Matsumoto and Hwang's (2013) judgment regarding the available evidence about the IDI specifically comes from their lack of reliance upon and citation of established criteria for evaluating measures. As such, their approach appears to be out-of-step with the existing literature on how to evaluate a measure.

One such available publication of criteria is the *Standards for Educational and Psychological Testing*, which now represents a 6th revision of previously published criteria.

The purpose of the Standards is to provide criteria for the development and evaluation of tests and testing practices and to provide guidelines for assessing validity of interpretation of test scores for the intended test uses ... the Standards provides a frame of reference to ensure that relevant issues are addressed. (p. 1)

Given the authors' conflict of interest, their lack of following established evaluation criteria, and not sufficiently citing sources to support their framework, questions arise as to the authors' selection of evaluation criteria that favored certain measures over others.

I appreciate Matsumoto and Hwang's (2013) honest appraisal about the marginal internal structure evidence for their own test (i.e., the ICAPS, p. 864). And yet, in their summary conclusion they favored the ICAPS over other measures. Without clearly articulated commonly accepted standards upon which to evaluate tests, it is difficult to discern why the CQ, ICAPS, and MPQ were included in the authors' summary interpretation and why the other measures evaluated were not included.

Central to the concern of disregarding long-established criteria is Matsumoto and Hwang's (2013) use of ecological validity as a distinct, primary evaluation criteria (p. 864); with the implication therefore that ecological validity was the standard upon which they arrived at their summary interpretation. The authors' equated ecological validity to the more commonly accepted terms/types of validity: criterion-related validity and external validity (Matsumoto & Hwang, 2013, p. 852). Their subsuming of criterion-related validity and external validity under ecological validity is certainly permissible, given that they defined their terms.

Yet, what remains unclear is why ecological validity was positioned as distinct from construct validity? The two most recent versions of the Standards (1999, 2014) reflect current psychometric thinking which posits construct validity as a unitary construct (see also, Cronbach, 1984; Cronbach & Meehl, 1955; John & Soto, 2007; Strauss & Smith, 2009). As Hoyt, Warbasse, and Chu (2006) stated, "there is now a well-established consensus that construct validation is the primary and unitary process and that content-related and criterion-related evidence are two important information sources for evaluating the construct validity of score interpretations" (p. 779). Why would Matsumoto and Hwang (2013) diverge from consensus and state "the main criterion against which a test of 3C should be evaluated is the strength of evidence for its ecological validity" (p 864)?

Further, Matsumoto and Hwang (2013) utilized a narrow definition of construct validity; equating construct validity to what is more commonly referred to as the internal structure of a measure (p. 864). Had the authors followed an established set of evaluation criteria such as the Standards, they would have seen that the internal structure of a measure is more commonly regarded as one potential source of information that can be used to form a judgment or validity argument about a measure's utility with respect to its intended purpose (AERA/APA/NCME, 1999, 2014).

Construct validation then subsumes content validity, and evidence of a measure's internal structure *and* external structure (Hoyt et al., 2006), with the latter referring to

“evidence based on relations to other variables” (AERA/APA/NCME, 2014, p. 16). In this framework, what Matsumoto and Hwang (2013) defined as ecological validity (i.e., criterion-related and external validity) is subsumed under external structure, along with convergent and discriminant evidence and generalizability evidence; with criterion-related evidence subdivided into concurrent and predictive empirical associations with other variables. Again, why would Matsumoto and Hwang (2013) diverge from consensus and limit construct validity to the internal structure of a measure? Construct validity is the chief process that content, internal structure and external structure evidence supports or limits.

A second concern related to Matsumoto and Hwang’s (2013) lack of adherence to commonly accepted standards for evaluating measures is that a judgment regarding a measure’s use is specific to “the *intended interpretations* of test scores and their relevance to the *proposed use*” advocated by the authors/developers of the specific test under review (AERA/APA/NCME, 2014, p. 11; italics added).

The type of work attempted by Matsumoto and Hwang (2013) is one that is inherently and inevitably incomplete. In their attempt to “provide a comprehensive review of the available tests of 3C” (p. 849), they chose to group all of the tests under a *global statement about intended purpose* (i.e., test scores are intended to assess an individual’s cross-cultural competence (Hoyt et al., 2006, p. 489). Evidence evaluated as supportive of or limiting a global statement about intended purpose does not address whether the evidence speaks to the specific stated claims test developers made about their own instrument.

The implication of the above understanding of validity is that the *evidence that Matsumoto and Hwang (2013) reviewed and formed judgments upon does not comprehensively and definitively speak to the specific intended uses of the IDI* as proposed by Hammer (2011; Hammer, Bennett, & Wiseman, 2003). According to Hammer (2011), the IDI is an assessment tool that measures the level of intercultural competence/sensitivity across a developmental continuum for individuals, groups, and organizations and represents a theoretically grounded measure of this capability toward observing cultural differences and commonalities and modifying behavior to cultural context. (p. 475)

Matsumoto and Hwang (2013) appear to understand that evaluative judgments about tests are inevitably highly idiosyncratic (i.e., validity arguments should be constructed specific to each separate test) when they acknowledged “the suitability of any test is relative to the specific domain of 3C that it was intended to predict and the [knowledge, skills, abilities, and other factors] that were theoretically hypothesized or empirically demonstrated to underlie the test” (pp. 867-868).

Yet, they concluded “that the CQ, ICAPS, and MPQ have the most promising evidence for assessing 3C” (p. 867). Do the authors mean, if a researcher wants to use a global assessment of cross-cultural competence, without respect to developmental theory or how cultural differences and similarities are perceived and then acted upon, then the CQ, ICAPS, and MPQ seem to warrant use for such global assessment? Again, not following commonly accepted standards for evaluation raises questions about the veracity of their interpretation.

A construct validation argument should answer the question “does the available evidence support the specific test for its intended use?” In other words, does the cumulative available evidence support the IDI as a measure that assesses different developmental capacities for (1) conceptualizing cultural differences and commonalities and (2) modifying behavior accordingly, as conceptualized by the Developmental Model of Intercultural Sensitivity? The IDI does not assess a broad, global definition of cross-cultural competence (i.e., “intercultural adaptation *and* adjustment,” Matsumoto & Hwang, 2013, p. 849).

The answer to the construct validity question is always conditional upon existing research to date, as well as each study sample, context, and time. The answer is always evolving and therefore incomplete at any point in time. This understanding permits the following statements by Matsumoto and Hwang (2013) to be made about any test, including the IDI: (1) further psychometric testing is needed, (2) there is a lack of test-retest reliability evidence reported in the existing literature, and (3) there is a need for research demonstrating incremental validity (p. 867). Such statements reflect current psychometric thinking about construct validation as a *process* and not a property of a test; an understanding that Matsumoto and Hwang (2013) appear to have given their use of language throughout their paper (e.g., “construct validity ... has been demonstrated” [p. 862] and “a number of studies have provided evidence for ...” [p. 862]).

And so, again, it is surprising that the authors offered no qualifiers or specific rationales for their judgment “that the CQ, ICAPS, and MPQ have the most promising evidence for assessing 3C” (p. 867). It seems that the authors were making a claim that the CQ, ICAPS, and MPQ have the most available cumulative evidence (i.e., the most published empirical studies) and therefore more evidence that supports their use as a global measure of 3C? *Their conclusion, however, says very little about the quality of the available evidence for the use of the IDI for its intended purpose.* Had the authors tapped into more of the available evidence for the IDI, and evaluated the IDI in terms of claims made about its intended use, then they likely would have reached a different conclusion.

Last, one inherent difficulty in test construction and validation is tautological problem (e.g., Koenig, 2008; Ng & Early, 2006; Skeem & Cooke, 2010). The tautological problem raises questions specifically about criterion-related evidence. Koenig (2008) described the tautological problem specific to research on religion and health and stated it this way,

This concept of spirituality includes not only indicators of traditional religiousness or a search for the sacred, but also positive psychological states: purpose and meaning in life, connectedness with others (quality of social support), peacefulness, harmony, and well-being. I call this the “tautological” version because by including indicators of mental health in the definition of spirituality, this assures a positive correlation between spirituality and mental health. Rather than conduct research to determine whether spirituality is related to mental health, this model solves the problem by defining spirituality from the very start in terms of mental health. Research that documents an association between spirituality defined this way and

positive mental health, is meaningless since constructs measured with the same or similar items will always be correlated with one another. (p. 350)

It appears that some of the research findings reviewed by Matsumoto and Hwang (2013), as evidence in support of the ecological validity for a particular test, could simply represent tautological associations. At a minimum, the question may be legitimately posed “what evidence is there that the criterion-related associations are not merely a product of tautology?”

When cross-cultural competence is operationalized by indicators of adjustment and then correlated with other indicators of adjustment, then Koenig’s (2008) conclusion seems applicable to the validation evidence for measures of 3C that assess cultural adjustment. In other words, what practical meaning is there in research that demonstrates an association between 3C adjustment measures and adjustment? Tautologies might elevate correlation coefficients used to document convergent and/or criterion-related validity, and yet, an adjustment-adjustment correlation tells us nothing of substance about 3C. The demonstrated association does not advance knowledge about 3C beyond an observation that adjustment in one domain corresponds to adjustment in another domain of functioning or experience.

Matsumoto and Hwang (2013) indicated that common domains or conceptual overlap existed among the CQ, ICAPS, and MPQ, and consisted of openness, flexibility, empathy, and emotion regulation; all which appear to be aspects of the broader construct of adjustment. Adjustment in the domain of cross-cultural encounters should correlate (i.e., demonstrate ecological validity) with adjustment more broadly defined, bringing into play the tautological problem. Hammer (2011) makes the claim that the IDI assesses 5 distinct worldviews (i.e., orientations, mental templates, or mindsets) specific to perceiving and interpreting cultural differences and similarities. The five worldviews fall along a continuum of cognitive-behavioral complexity, with the worldview of adaptation defined as “an orientation that is capable of shifting cultural perspective and changing behavior in culturally appropriate and authentic ways” (p. 475).

In terms of the intended use of test scores specific to the IDI, the IDI *does not assess adjustment*. According to Matsumoto and Hwang, adjustment refers to “subjective experiences associated with adaptation” (p. 850). They go on to say that adjustment “may be assessed by mood states, self-esteem, self-awareness, physical health, self-confidence, stress, psychological and psychosomatic concerns ...” (p. 850). The difficulty for measures of cross-cultural competence that attempt to assess 3C adjustment is the question posed by Ng and Early (2006), commenting specifically on the measurement of CQ, is the measure “an indicator or a predictor of cross-cultural effectiveness?” (p. 8).

The IDI distinguishes between adaptation and adjustment and makes no claim to assess subjective (i.e., psychological) experiences when assessing 3C, and therefore has the advantage of being able to more clearly distinguish predictor (i.e., capability, adaptation, sensitivity) from outcome (i.e., adjustment, subjective experiences, performance in a distinct domain). In other words, in order for criterion-related evidence to be most

meaningful, the predictor and outcome should represent conceptually distinct domains. Intercultural adaptation is distinct from adjustment. Matsumoto and Hwang draw this conceptual distinction, and the distinction exists elsewhere in the literature.

Since the IDI focuses on assessing adaptation, when the IDI demonstrates associations with adjustment outcomes the associations do not stem from a tautology. I am not sure this can be said about all of the 3C measures, and especially perhaps the ICAPS given Matsumoto and Hwang's approach to test construction (p. 852; i.e., "removing items not associated with criterion variables"). If a 3C test consists of items operationalizing adjustment, then demonstrated empirical associations with outcome measures of adjustment, is this not indicative of a tautological problem? The CQ and the IDI are more immune from tautological associations because they measure cognitive capacities and behavioral skills specific to adaptation. And when utilized to predict outcomes the empirical associations can more clearly advance knowledge about what can happen when adaptation occurs.

As a reader of Matsumoto and Hwang's Article, with a particular understanding of construct validation, I offered my observations to temper or qualify their judgment. I also offer these observations in the spirit of clarification. Research begins with a carefully constructed question, and when designing a study to answer that question, including the decision about which measure to use to operationalize one of their variables, 3C researchers must consider the available construct validation evidence for the different measures that exist. My hope is that the observations I have offered provide the reader of Matsumoto and Hwang's (2013) article with an additional, more nuanced understanding of the choice about instrumentation.

Inaccurate Statements Regarding the IDI

I now turn to a delineation of concerns regarding inaccurate statements made by Matsumoto and Hwang (2013) in their characterization of the IDI. For reference purposes, quotes/information are taken directly from the Matsumoto and Hwang's review Article and also from two IDI validation articles. The initial IDI validation article was published in 2003 (Hammer 2003). The second IDI validation article was published in 2011 (Hammer 2011). This follow-up validation article (Hammer 2011) provides a detailed summary of prior validation research and also presents additional validation studies conducted since 2003. Below then is a summary of inaccuracies specifically regarding the treatment of the IDI by Matsumoto and Hwang (2013).

1. *Matsumoto and Hwang state: "Hammer and colleagues developed the IDI by first interviewing 40 international student volunteers . . ." (p. 860).*

Interviews were conducted with individuals who were explicitly not international student volunteers; they were working individuals across the age and educational spectrum. Given previous critiques of instrument development vis-à-vis item identification as relying too heavily only on student respondents, the IDI explicitly developed the item

pool from a much wider range of respondents beyond the “student volunteer.” This was made clear in Hammer (2003):

“It was decided that the actual sample of interviewees would consist of people of varied cultural background 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 places of employment in and around the Washington D.C. area . . . Overall, the sample of 40 individuals represented a culturally diverse group with varying international experiences” (p. 427).

The above provides evidence of content validity for the IDI, as part of an overall construct validity argument. In addition, using Matsumoto and Hwang’s definition of ecological validity, or the more commonly accepted notion of generalization, the sample utilized provides evidence of cultural generalizability (Hoyt et al., 2006).

2. *It is not accurate for Matsumoto and Hwang to conclude: “Worldmindedness and intercultural anxiety correlated with three of the five scales, providing some evidence for construct and ecological validity” (p. 860).*

Five correlational hypotheses were proposed in Hammer (2003). Three of the hypotheses proposed significant correlations while two hypotheses predicted nonsignificant correlations. In fact, all hypothesized correlational results were in the predicted direction. It is disingenuous to state that because three of the five IDI scales correlated with Worldmindedness and Intercultural Anxiety, this provides “some” evidence of construct and ecological validity. In fact, given that the correlational analyses were in the hypothesized direction, as informed by theory (i.e., the Developmental Model of Intercultural Sensitivity; which is the theoretical context for the IDI), the associations provide evidence of construct validity (see Hammer, 2003, pp. 437-438).

Also, correlations with a social desirability index showed no significant correlations; indicating that individuals do not “figure out” how to score higher on the IDI. Social desirability is a confound that introduces construct irrelevant variance when examining associations between measures (Hoyt et al., 2006). Demonstrated correlations between measures could be due to a social desirability confound rather than actual “substantive [associations] between the focal constructs” (Hoyt et al., 2006, p. 786). This is an important piece of information about the IDI which offers further construct validation evidence; a finding omitted from Matsumoto and Hwang’s review of the IDI.

3. *Matsumoto and Hwang state on page 860 that “A few studies have provided evidence of the ecological validity of the IDI with demographic variables . . . “. And later, on page 866: “The only evidence for the ecological validity of the IDI comes from studies reporting significant correlations between it and demographic variables.”*

Matsumoto and Hwang omit two predictive validation studies discussed in Hammer (2011: pp. 484-485) that provide criterion-related evidence for the construct validity of the

IDI. In one study, a total of 71 recruiters of the U.S. staffing function within a high tech, multinational organization participated. They worked in six teams, located throughout the United States (team 1, n = 15; team 2, n = 8; team 3, n = 13; team 4, n = 10; team 5, n = 13; and team 6, n = 12). The correlations between the team IDI Developmental Orientation scores and separately, individual IDI results of each member of the six teams to the degree to which cultural diversity was reflected in each of the six team's U.S. hiring results was examined. External benchmarking data was computed from U.S. government workforce data by hiring position. Based on this external benchmarking, the number of diversity hiring benchmarks missed by each team was calculated (missing of a benchmark was indicated by missing the benchmark by at least 20%). Two separate correlational analyses were done: Team IDI scores and number of diversity hiring benchmarks missed by team, and individual IDI scores and number of benchmarks missed. Results indicated very strong negative correlation between overall team IDI scores and team benchmarks missed. Pearson product-moment correlations (Pearson's r) run between individual IDI Developmental Orientation scores (n = 71) and each individual's overall team Diversity goal achievement rating again resulted in strong, significant negative correlations. Overall, lower scores on the IDI (team and individual) were predictive of more diversity benchmarks missed. This is a real world study of diversity hiring patterns within a large corporation conducted over a year-long period. This offers strong support of criterion-related validity, which is part of the larger validity argument for the IDI—yet is not reported by Matsumoto and Hwang in their review.

In another study (Hammer 2011), also not reported by Matsumoto and Hwang, the relationship between IDI scores and study abroad outcomes were examined. 1,500 study abroad students and 638 control group (home campus) students from Austria, Brazil, Costa Rica, Ecuador, Germany, Hong Kong, Italy, Japan and U.S. participated. Significant correlations between (higher) IDI scores and study abroad outcomes in the areas of increased knowledge of the host culture, less intercultural anxiety, increased percentage of intercultural friends (back in home campus) and higher levels intercultural satisfaction were found (p. 485).

These demonstrated associations provide further evidence of criterion-related validity, again as part of a larger construct validation argument for the IDI. The validity argument is strengthened by this evidence because the IDI avoids the tautological problem and assesses adaptation as a distinct construct from that of adjustment. Consistent with Ng and Early's (2006) assertion about the CQ when referencing the tautological problem, the IDI measures adaptation as "a capability that is posited to *predict, but is distinct from*, the actual outcome arising from a specific situation or episode of interaction" (p. 8; italics added). These findings suggest that adaptation (i.e., increased cognitive complexity in perceiving and behaviorally responding to cultural differences and similarities) corresponds to gains in adjustment outcomes. These demonstrated associations are not due to construct irrelevant variance because the items on the IDI are specific to adaptation as conceptualized by the DMIS and are not based on items derived from indicators of adjustment (i.e., the items are from a distinct content domain than the outcomes).

4. *Matsumoto and Hwang erroneously claim that "pre-post tests using the IDI to assess training*

efficacy, however, have produced mixed results . . .” (p. 860).

The theoretical model of the IDI overtly suggests that training and education efforts that are targeted developmentally for the primary orientation of the individual can lead to gains in intercultural competence (that is, higher posttest scores); however, interventions that are not developmentally targeted based on IDI results would not necessarily be expected to increase an individual’s intercultural competence (no significant differences in pre/posttests). The fact that pre/posttest studies produced mixed results in terms of gain scores is theoretically not surprising and completely unrelated to the validation of the IDI. This again highlights the need for evaluations of measures to be specific to each specific test, and the stated intended purpose for a particular test. Construct validation evidence must be examined within the specific theoretical context of the measure and its intended uses. Further, evidence presented in the book by M. Vande Berg, R.M. Paige & K. Lou, (2012) *Students Learning Abroad* support the observation that when training is developmentally targeted based on IDI results, significant gains in intercultural competence as measured by the IDI are achieved and when interventions are not developmentally designed based on IDI results, often little gain in intercultural competence occurs. These findings therefore offer further construct validation evidence (i.e., criterion-related evidence) that the IDI measures intercultural competence as the theory in which the measure is embedded would predict.

5. *Matsumoto and Hwang acknowledge on page 860 that content validity was tested in developing the IDI through: “A panel of intercultural experts reviewed the items pool; items that could not be reliably categorized by an agreement rating of .60 or higher were eliminated, resulting in a 145-item pool” (p. 860).*

While Matsumoto and Hwang describe this content validation protocol on page 860 in their review Article, Matsumoto and Hwang omit this protocol in their subsequent analysis of content validity of the IDI later in their Article on page 864 under “Content Validity”.

6. *Matsumoto and Hwang ignore in their review of the IDI that the IDI items were generated in a way that addresses cultural bias, a particular validation strength of the IDI.*

A critical protocol for instrument validation involves identification of protocols to address possible cultural bias in the way items are generated. The IDI items were generated in a way that addressed cultural bias. Specifically, from the full 40 transcribed interviews, actual statements (from a culturally diverse set of interviewees) made by the interviewee’s associated with their rated DMIS orientation were identified and added into the item pool. This item pool was then reviewed by two, culturally diverse pilot groups for item clarity, redundancy etc. As Hammer (2011) explains:

“Because the “data set” was verbatim statements made by the interviewees, the actual statements made by this culturally diverse group were identified for each of the developmental orientations. Thus, unlike other instrument development approaches, the actual items of the IDI were originally generated in natural discourse by people from a wide range of cultures. This is in contrast to questionnaire item that are

generated by the researchers themselves. This original innovative approach to generating the item pool provides evidence for the content validity of the items and counters possible criticism of systematic cultural “bias” being introduced by the researchers in the wording of the items themselves” p. 476).

The item construction process utilized in the development of the IDI provides evidence of cultural generalizability.

7. *Matsumoto and Hwang state regarding the IDI that “. . . assessments of the factor structure of the IDI have produced inconsistent results in limited samples that do not correspond to the model intended to be measured. Nonconfirmation of structural validity renders correlations of scale scores to other constructs questionable because it is not clear what the scales are assessing.” (p. 864).*

Across all study samples, a total of 10,000 respondents from a wide-range of cultures throughout the world validated the IDI. This is not what one would characterize as “limited samples.” The development of the IDI was done through an iterative process that indicates the confirmatory factor analysis results reviewed in Hammer (2011) provide persuasive evidence of the structural stability of the IDI across culturally diverse groups. This is elaborated below.

Study 1: Factor analysis of 123 items with a diverse sample of 226 subjects. Results identified: Denial, Defense, Minimization, Acceptance, Cognitive Adaptation and Behavioral Adaptation scales. Reversal and Integration orientations did not emerge as stable factors. As an initial test of a complex theory, these results were encouraging.

Study 2: Paige (2003) factor analyzed the 60 item IDI and identified six factors: Denial/Defense, Physical Universalism (a sub-construct of Minimization), Transcendent Universalism (the second sub-construct of Minimization), Cognitive Adaptation, Behavioral Adaptation and a sixth factor not identified. Overall, their results generally confirmed the theoretical constructs; they did find that Physical and Transcendent Universalism clustered between the more ethnocentric orientations of Denial/Defense and the more ethnorelative orientations of Cognitive and Behavioral Adaptation (the two sub-constructs specified by the DMIS theory). This finding does not show “inconsistency” in the IDI; rather, it suggested possible modification of the DMIS theory, which is an inevitable element of the construct validation process (Hoyt et al., 2006). The theory of DMIS characterizes Minimization as ethnocentric while Paige et al (2003) suggested it may be more of a transitional mindset. Nevertheless, the Paige (2003) study supports the placement of the two Minimization scales on the developmental continuum as following Denial and Defense and before Acceptance and Adaptation.

It is important to again note that commonly accepted standards for evaluating construct validity indicate that construct validation is a process that occurs in the context of an evolving theory (Hoyt et al., 2006). This is particularly relevant in the case of the IDI, which is theoretically grounded in the DMIS. According to Strauss and Smith (2009), construct validation “is an iterative process ... [that] provides information that leads to theory

refinement and elaboration, which in turn provides a sounder basis for subsequent construct and theory validation research” (p. 9). Strauss and Smith refer to this as an “informative approach” that unites theory testing and construct validation. Matsumoto and Hwang’s (2013) statements about the “inconsistency” of the internal structure evidence about the IDI seems out of step with current thinking about construct validation. Inconsistency should be more precisely seen as theory and measurement refinement.

Study 3: Hammer (2003) reviewed the original set of 145 items in the IDI and added some items to better reflect the two DMIS constructs that were to date, not found as stable factors: Reversal and Integration. These items were administered to 591 culturally diverse respondents along with the Worldmindedness scale and an intercultural anxiety measure and a social desirability measure. In this case, confirmatory factor analysis (CFA) was used. These results revealed that BOTH a seven dimension model (Denial, Defense, Reversal, Minimization, Acceptance, Adaptation and Encapsulated Marginality) and a five dimension model (Denial/Defense, Reversal, Minimization, Acceptance/Adaptation and Encapsulated Marginality) were an equally good fit to the data. While the seven dimension model fit the data equally well as the five factor model, “based on the criteria of parsimony, the five-factor model was selected” (Hammer 2011 p. 477). Further, as mentioned above, criterion-related validity testing of the IDI scales yielded associations to Worldmindedness and intercultural anxiety that were in the predicted direction.

Study 4: The IDI was administered to a culturally diverse sample of 766 respondents in order to develop a “total IDI score” that would represent an individual’s overall orientation toward cultural differences along the theoretical continuum. In addition, a Perceived total score was also computed, with both overall scales achieving reliabilities of .83 and .82, respectively.

Study 5: The IDI was administered to 4,763 culturally diverse respondents from eleven distinct cross-cultural samples. Confirmatory Factor Analysis was conducted. Because the seven factor model AND the earlier selected five factor model had equally good fit to the data (recall the five factor model was then selected due to parsimony; not because it fit the data better); therefore, study 5 was conducted to test these two models with a larger sample. In this study, the seven factor model did emerge as a better fit to the data compared to the five factor model; (the seven factor model more closely mirrors the DMIS model). This was tested both across the full sample (aggregate data) and for each individual, culturally distinct cohort (a culture specific test protocol). In each case, the seven factor model was the best fit to the data. Scale reliabilities were again found to be robust. As Hammer (2011) explains:

“While some modifications to the original DMIS orientations arise as a result of the collective Phase 1, 2, post Phase 2 testing, and the current Phase 3 validation studies of the IDI, an overall conclusion from these various efforts is that the main theoretical insights offered by the DMIS are consistently confirmed. These studies provide overall support for the DMIS as a fundamentally sound theoretical framework and simultaneously support the modifications to the DMIS framework presented in this article” (p. 482).

Construct validation research should evolve theory. The fact that empirical research operationalizing propositions from the theory of DMIS resulted in modifications to the DMIS framework is non-problematic from the perspective of current approaches to construct validation. *Taken together, the internal structure evidence across multiple study samples demonstrate that Denial, Defense, Minimization, Acceptance, and Adaptation emerge as stable yet refined dimensions of intercultural competence as defined by the DMIS and informed by results of empirical research on the IDI.* The fact that the IDI itself evolved from 60-item to 50-item measure further illustrates the dynamic iterative process of theory testing and construct validation.

Study 6: Correlations among the seven dimensions were assessed in the sample of 4,763 subjects. As stated in Hammer (2011):

“Noteworthy about this correlation matrix are the following. First, there is a strong correlation between Defense and Denial ($r = .83$) and a strong correlation between Acceptance and Adaptation ($r = .64$). Second, Reversal is positively correlated with Denial (.36) and with Defense (.37) and not substantially correlated with Acceptance (.01) or Adaptation (.12), consistent with the finding and explanation from the initial validation study (Hammer 2003). Third, the very weak correlations of Minimization with both the more monocultural orientations (Denial, Defense, Reversal) and the more intercultural orientations (Acceptance, Adaptation), suggest Minimization exists as a transitional orientation between the more monocultural and intercultural orientations. As a transitional orientation, Minimization is thus more accurately conceptualized as being not ethnocentric (monocultural) yet also not as interculturally competent/sensitive as Acceptance or Adaptation. Fourth, there are negative correlations between Defense and Denial scales and the Acceptance and Adaptation scales which provide support for the developmental continuum.” (p. 481).

The above results demonstrate evidence of convergent and discriminant validity in the context of examining the internal structure of the IDI, as associations between dimensions corresponded to convergent associations among theoretically similar constructs (i.e., convergence among ethnocentric dimensions and convergence among ethnorelative dimensions) and divergent associations among the scales of theoretically dissimilar constructs (i.e., ethnocentric scales diverge from ethnorelative scales).

It is important to point out that the evidence reviewed by Matsumoto and Hwang is incomplete, and yet despite this the evidence in their review can be interpreted, in light of contemporary thinking about construct validation, as evidence supporting the use of the IDI as a measure of intercultural competence. And specifically, a measure of intercultural competence informed by the DMIS which posits distinct worldviews (i.e., increased cognitive complexity in perceiving and behaviorally responding to cultural differences and similarities) at different stages of intercultural development.

Conclusion

The information presented in this reply render the Matsumoto and Hwang (2013) evaluation of the IDI quite problematical. Unfortunately, their Article has already appeared in JCCP. It is my hope that readers will not simply accept at face value Matsumoto and Hwang's evaluation of the IDI without taking into consideration the information presented here and will verify for themselves the veracity of the claims made by Matsumoto and Hwang.

References

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.
- Hammer, M. R. (2011). Additional cross-cultural validity testing of the Intercultural Development Inventory. *International Journal of Intercultural Relations*, *35*, 474-487. doi:10.1016/j.ijintrel.2011.02.014
- Hammer, M. R., Bennett, M. J., & Wiseman, R. (2003). Measuring intercultural sensitivity: The Intercultural Development Inventory. *International Journal of Intercultural Relations*, *27*, 421-443. doi:10.1016/S0147-1767(03)00032-4
- Hoyt, W. T., Warbasse, R. E., & Chu, E. Y. (2006). Construct validation in counseling psychology research. *The Counseling Psychologist*, *34*, 769-805. doi:10.1177/0011000006287389.
- Jankowski, P.J. (under review). The Intercultural Development Inventory: A construct validation argument.
- John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, R. F. Krueger, R. W. Robins, R. C. Fraley, R. F. Krueger (Eds.) , *Handbook of research methods in personality psychology* (pp. 461-494). New York: Guilford Press.
- Koenig, H. G. (2008). Concerns about measuring 'spirituality' in research. *Journal of Nervous and Mental Disease*, *196*, 349-355. doi:10.1097/NMD.0b013e31816ff796.
- Matsumoto, D., & Hwang, H. C. (2013). Assessing cross-cultural competence: A review of

Available tests. *Journal of Cross-Cultural Psychology*, 44, 849-873.
doi:10.1177/0022022113492891.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Ng, K. Y., & Earley, P. C. (2006). Culture+intelligence: Old constructs, new frontiers. *Group & Organization Management*, 31, 4-19.

Skeem, J. L., & Cooke, D. J. (2010). One measure does not a construct make: Directions toward reinvigorating psychopathy research—reply to Hare and Neumann (2010). *Psychological Assessment*, 22, 455-459. doi:10.1037/a0014862.

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5, 1-25.
doi:10.1146/annurev.clinpsy.032408.153639.