JOURNAL OF NARRATIVE AND LANGUAGE STUDIES | ISSN: 2148-4066
NALANS
*Journal of Narrative and Language Studies – December 2019, Volume 7 – Issue 13*

# Intercultural Competence Assessment Formats: Reliability and Validity Formats

*Elaheh Rahimi*[1]

*Kerman Institute of Higher Education*
*elaherahimi496@yahoo.com*

**APA Citation:**
Rahimi, E. (2019). Intercultural Competence Assessment Formats: Reliability and Validity Formats. *Journal of Narrative and Language Studies, 7*(13), 221-258.

**Abstract**

Globalization has encouraged a demand for increased intercultural competence (ICC) in higher education. Although in the modern world, the assessments and concepts of ICC have been given a wide attention, few assessments have designed to meet the standards in areas of innovative item types and reliability and validity evidence. This study seeks to discuss the possible item types and their strengths and weaknesses within the category of selected response items. In addition, this paper aims to discuss the reliability evidence for the previously designed ICC assessments and then discuss the validity evidence concerning the internal structure, the relationship with conceptually related constructs, and the relationship with criteria. To summarize, the researcher found that the reliability evidence of existing assessments includes no major issue with reliability at the total test level. Regarding validity, it was found that for most assessments, insufficient validity exists, particularly criterion-related validity.

**Keywords:** Intercultural competence, assessment format, reliability, validity evidence

## 1. Introduction

In today's world educators and employers considerably affirm the importance of intercultural competence. Although the majority of higher education institutions support the idea that these skills provide valuable results for their graduates, few of these institutions have adequate means by which they can assess the wide diversity of outcomes. Having and using intercultural assessments will encourage and support researchers to perceive and assess the efficiency and results of their respondents and assist them in developing correct interventions and answers at different levels as well. Several assessments exist that are particularly developed for the assessment of ICC.

At this time surveys and portfolio assessments are two important assessments formats to measure ICC. The instruments presented in table 1, are surveys which range from one item (i.e., Global perspective survey; Hanvey, 1982) to over 160 items (i.e., Intercultural Communication

[1] Postal address: Yasuj. Ebnesina Street 11. Plaque. 14.. Iran. Phone number: +989172283186

and Collaboration Appraisal; Messner & Schafer, 2012). Generally, the format of this kind of assessment is online, although some assessments (e.g., Intercultural Development Inventory; Hammer, Bennet, & Wiseman, 2003) are delivered in paper and pencil format. In this paper, the author only examined ICC assessments that exclusively used selected-response items.
Like surveys, portfolios also consist of constructed-response items. It is considered to be one of the most influential instruments to assess ICC in higher education. Typically, a portfolio assessment defined as a set of material which created by a person progressively or scores achieved from different assessments. At present, there are no standard portfolio assessments because the scoring method, content, and platform differs over institutions, contexts (e.g., general education, study abroad, experiences and foreign languages), and studies (e.g., Ingulsrud, Kai, Kadowaki, Kurobane, & Shiobara, 2002; Jacobson, Sleicher, & Maureen, 1999). This deficiency could be regarded as a benefit. Capturing the skill which are context-specific and developing of these skills gradually are two predominant jobs of portfolios; hence, it is believed that students could capture ICC via a set of work products in different times (e.g., before, during, and after an experience in a foreign country; Ingulsruds et al., 2002; Jacobson et al., 1999).

Digital portfolios are used by a number of higher education institutions across the world. For instance, a digital portfolio issued in Clemson University and students were required to prepare evidence of cross-cultural awareness as a vital part of education at the university. Evidence of cross-cultural awareness is represented in digital portfolios via the involvement of samples of writing. Alliant International University uses a digital format for assessment of ICC in its study abroad students. in spite of the fact that portfolios also have this potentiality to encompass other work products like intercultural communication's audio and video recording, no institutions have been identified to request for such products.

The format of all assessments considerably depends on the intelligent goal of the assessment. Although, it is suggested that the researchers could use more than one methodology (i.e., both quantitative and qualitative methods) to assess ICC ( Deardoff, 2006; Fantini, 2009), measuring students' ICC in higher education institutions need a holistic format which provides a comprehensible comparisons of individuals and groups of examinees. According to ICC experts, portfolios may not be an appropriate format for assessing ICC, because it is to some extent difficult to standardize the different work products offered by students and to guarantee inter-rater reliability in scoring the work products by students. However, a survey is more advantageous, as it is more standardized and norm-referenced and allows higher education institutions to make a presumption about an individual and a group of ICC.
Furthermore, surveys consist of diverse kinds of selected-response in term formats which represent the multidimensional nature of ICC more clearly. For instance, Likert-scale responses are more appropriate to assess the attitudinal components of ICC, but forced-choice or multiple choice questions are more adequate to capture the skills and knowledge that describe ICC.

In this paper, first, I discuss the various item format and their strengths and weaknesses within the group of selected-response items, and then, the reliability and validity evidence related to these assessments will be discussed.

## 2. Selected Response Items

### 2.1 Multiple-Choice Items

Multiple choice items are used to assess the knowledge component of ICC such as in the global competence Aptitude Assessment (W. D. Hunter et al., 2006) and Global Awareness Profile (GAP; Corbiti, 1998). There are distinctions among these assessments, as some multiple choice items capture knowledge which is particular to one culture and other assesses cultural knowledge that is general or worldwide. Like culture-general knowledge, the multiple-choice items are used by the GAP to measure knowledge of politics, religion, geography, environment, and socioeconomics of six regions (Africa, South America, North America, Asia, the Middle East, and Europe) across the globe. On the contrary, the global competence Aptitude Assessment (Global Leadership Excellence, 2010) applies multiple-choice items according to specific culture regardless of the inclusion of any culture-general items.

## 2.2 Likert-Scale Items

Likert-scale items commonly include the statements which range from a strongly agree to strongly disagree. Typically, some assessments directly ask respondents to measure themselves on a specific skill. For example, a behavioral regulation item is likely to ask respondents to show if they change their behavior according to cultural customs. The next alternation akin to ICC assessments with Likert items is the number of response points or categories on the response scale. A large number of assessments exert a 5-point Likert scale, whereas others vary from a 4-point to 7-point scale. Most Likert-type items are self-report, however, one assessment that introduced in this review exploited Likert-type responses for peer assessment. One instrument that uses a 4-point Likert scale is a peer rating of intercultural communication effectiveness is the Behavioral Assessment Scale. This assessment scale was extracted from Ruben's (1976) behavioral assessment of communication competency for intercultural adaptation (see Chen, 1992, for a review). This instrument was applied for two roommates. One is an international student and the other is a native speaker of United State. Both roommates assess each other based on eight items which assess the following features of ICC: empathy, interaction posture, relational roles, display of respect, tolerance for ambiguity, and interaction management. Each one item scale, unlike the other ICC assessments, demonstrates the behavioral description of the roommates that they are rating for each of the 4points on the Liker Scale. The only assessment which consists of this kind of description for Liker Scale anchor is the BASIC because a huge number of assessments potentially exploit more traditional Likert scale response categories (i.e., strongly agree, to strongly disagree).

## 2.3. Implicit Association Tests and Q-Sort Methodology

Implicit Association Tests (IATs) and Q-sort methodology are item formats that are less common to measure the attitudinal component of ICC. IATs main roles are to assess how strongly a respondent could connect two mental concepts, or representations, by assessing the response time for making the correct association (Greenwald, Poehiman, Uhlmann, & Banaji, 2009). Typically it means that if respondent relates an object to a concept faster he/she can perceive that there is a stronger relationship between those mental those mental concepts. One kind of IAT is Test of Hidden Bias that measure negative prejudices toward different ethnic groups. For example, examinees are demonstrated with two images of an Africa American face next to a White face on a computer screen and asking the respondents to quickly opt the "bad" or "good" photo. As in this case there is no correct association, per se, the authors state that "faster responses for the (black + positive) (white + negative) task than for the (white + positive)

(black + negative) task indicate a stronger association of black than of white with positive valence" (Greenland et al., 2009, p.18).

Because this kind of assessments are specific to America's context where the concept of race conceptualized as ethnically dichotomous in terms of white vs. black, IATs have been criticized by many. Another method which has been used ICC assessments is Q-Sort. The Q-Sort methodology plays an important role in psychology field of study and includes ran ordering of subjective concepts. The Q-Sort methodology is exploited by the Intercultural Communication and Collaboration Appraisal instrument (ICCA) designed by Messner and Shafer (2012). This tool asks individuals to arrange cards in response to a stimulus. The ICC consists of two Q-Sorts. The first kind includes the examinee arranging forty-eight behaviors, attitudes/ and belief in order from the most descriptive of self to least descriptive. In the second sort, the examinee must select the most predominant six intercultural competencies from a set of twelve competencies and arrange them from the less important to the most important one.

## 2.4 Situational Judgment Test

Situational Judgment Test (SJT) is another method of measuring ICC. One important goal of SJT is to assess a competing or ability according to the selection of responses by examinees in a hypothetical situation. Before selecting the pertinent response option of the presented set or respond to an open-ended stimulus, participants are required to read a few sentences of a real situation. The majority of the SJT stimulus gives importance to knowledge and behavioral components. For example, "what would you do?" is a prompt that requires examinees to represent the behavior that they would most likely to engage in from a number of possible actions. (Whetzel & McDaniel, 2009). The options are scored based on most effective, neutral, and ineffective behavior to make a composite score for the SJT. Knowledge stimulus such as "what is the best answer?" asks the examinees to select the appropriate response in the given situation. It is important that participants arrange the answers in order of most effective to least effective (Whetzel & McDaniel, 2009). As current meta-analysis shows, SJTs represent the remarkable content, criterion, and face validity (Whetzel & McDaniel, 2009).

However, SJTs typically have low internal consistency, as represented by Cronbach's alpha, due to the multidimensional nature of many SJT items. The internal consistency has been indicated by Cronbach's alpha. Base on this reason, the use of test-retest reliability or parallel forms are recommended by experts as testing SJT items' reliability in place of applying Cronbach's alpha (Whetzel & McDaniel, 2009). The "correct" response option has this potential to bias the test. If examiners are not aware of their cultural assumptions, this method is open to bias for cross-cultural SJTs. Participants show positive views toward this type of test (Lievens, Peeters, Schollaert, 2008). In addition, by measuring intentions, this type of test is more appropriate to measure behavior and skill rather than attitudinal measures.

There are a few examples of SJTs that are related to context of ICC, as though the critical incident format used in SJT items is found in cultural assimilators such as cross cultural training courses in which participants are presented with alternative behavioral options and cultural scenarios which they can discuss (Bhawuk, 2001; Earley & Peterson, 2004). For the Cultural Intelligent Assessment, respondents are required to select among a class of behaviors to show what option is the most appropriate for a given scenario (Thomas et al., 2015). Examinees must

answer 14 questions developed to assess cultural knowledge, metacognition, and skills. The next SJT has been developed to assess Cross-Cultural Social Intelligence (CCSI; Ascalon, Schleicher & Born, 2008). After responding to a series of cross-cultural scenarios, participants are asked to evaluate the likelihood that they would perform each of four behavioral choices. These four options are fallen into particular categories ( nonempathetic, nonethnocentric, nonempathetic, ethnocentric, empathetic, nonethnocentric, and empathetic, ethnocentric), taking into consideration the generation of two subscales: empathy (α□=□.61) and ethnocentrism (α□=□.71). Coefficient alpha for the overall scale was α□ = □. 68 (Ascalon et al., 2008). CCSJ as an example of SJT measure could represent the relationships with personality constructs (Ascalon et al., 2008) and cognitive ability (e,g., GMAT; r □=□.30). It has been indicating that GMAT has sufficient reliability (α□=□.92 for the test as a whole), particularly, the relationship between the three of Goldberg's (1999) International Personality Item Pool (IPIP) sub-dimensions (openness to experience, emotional stability, and consciousness) and the CCSI scores averaged r□=□.30. The IPIP represents sufficient overall internal reliability (α□=□.80). The CCSI has low reliability (α□=□.68 for the overall, α□=□.61 empathy subscale, and α□=□. 71 for the ethnocentrism subscale). However, these coefficients are almost the same as other SJT studies (Chan & Schmitt, 1997). The combination of convergent validity and internal consistency were considered as a strong indicator of the initial validity of both the use of SJTs and the measure to assess ICC. Nevertheless, there is no SJT specific to ICC that represent the evidence of criterion validity (Ascalon et.al., 2008).

## 2.5 Simulation-Based Measurement

Simulation-Based measurement is another commonly used training tool for measuring ICC. (e.g., Harrison, 1992; Jarrell, Alpers, Brown, & Wortring, 2008). Role-playing activities are more typical in simulations in which individuals take part in a limited intercultural scenario. The examinees are required to interact with an avatar (a figure representing a person or a computer-simulated character) or a confederate ( a paid assistant who has been instructed to act in a special way) who might demonstrate the cultural norms of a different group, his or her own cultural norms, and fictitious norms. Depending on the simulation, other individuals in the simulation could play this role in place of confederates. It is believed that one of the commonly conducted and popular simulation is the BaFa ' BaFa' simulation (Shirts, 1971). In this instrument, students are asked to imagine they are in two different fictional cultures and have interaction with each other for gathering a certain number of cards.

The two cultures are loosely developed to separate individual-collectivism diversity (prefer for group vs. individual) with verbal and nonverbal differences included (i.e., preferences for volume and personal space). Apart from the achievement of the goals of the game, individuals who watch could collect interaction data to measure the behavioral component of ICC. It is necessary to validate this measure; however, the present simulation kit doesn't have any behavioral checklist. The simulation by Harrison (1992) is a psychometrical one. This situation asks participants to interact with a confederate who pretend to manage a Japanese employee. The two rates the interaction independently based on soliciting employee input, maintaining harmony, improving consensus, demonstrating personal concern, and reducing conflict (Bhawuk & Brislin, 2000). Robin Sage Exercise is another well-known simulator (Skinner, 2002).

**Table 1**. Existing Assessments of Cross-Cultural Competence

| Test | Developed (year) | Format | Delivery | Forms and items | Themes/topics |
|---|---|---|---|---|---|
| Cross-Cultural Adaptability Inventory (CCAI) | Kelley and Meyers (**1995**) | Self-report; 5-point Likert scale (*definitely not true* to *definitely true*) | Paper and pencil/ Online survey | 50 items (4 subscales; 7–18 items per scale) | Emotional resilience, flexibility/openness, perceptual acuity and personal autonomy |
| The Global Perspective Survey | Hanvey (**1982**) | Self-report; 5-point Likert scale (*strongly agree* to *strongly disagree*) | Online survey | 9 items | Process of cross-cultural relativism in which one is able to view his/her own culture in relation to other cultures while suspending judgment and ethnocentrism |
| Assessment of Intercultural Competence (AIC) | Fantini and Tirmizi (**2006**) | Self-report; 6-point Likert scale (*not at all competent* to *extremely high competence*) | Online survey | 54 items (4 subscales; 11–19 items per scale) | Includes four dimensions: knowledge, attitudes, skills, and critical awareness |
| Intercultural Adjustment Potential Scale (ICAPS) | Matsumoto et al. (**2001**) | Self-report; 7-point Likert scale; anchors unknown | Online survey | 55 items | Measures cross-cultural competence through four psychological skills: emotional regulation, openness, flexibility, and critical thinking. |

| | | | | | |
|---|---|---|---|---|---|
| Cultural Intelligence Scale (CQS) | Ang et al. (**2007**) | Self-report; 7-point Likert scale (*strongly disagree* to *strongly agree*) | Online survey | 20 items | Measures cultural intelligence through four subscales: cognitive (knowledge of other cultures), metacognitive (awareness of how one thinks about other cultures), behavioral (behaving appropriately in cross-cultural interactions), and motivational (desire to interact with and learn more about other cultures |
| Global Competencies Inventory (GCI) | Bird et al. (**2002**) | Unknown | Online survey | 159 items | Measures leadership competencies of corporate managers and global leaders in areas critical to interacting and working effectively with people from different cultures. |
| Intercultural Development Inventory (IDI) | Hammer (**2011**) and Hammer et al. (**2003**) | Self-report (with 10 additional demographic items); 5-point Likert scale (*disagree* to *agree*) | Online and paper and pencil | 50 items | Measures orientations to cultural differences through five dimensions: denial/defense, reversal, minimization, acceptance/adapt |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | ation, and encapsulated marginality |
| Intercultural Sensitivity Scale (ISS) | Chen and Starosta (**2000**) | Self-report; 5-point Likert scale (*strongly disagree* to *strongly agree*) | Online survey | 24 items | Measures intercultural sensitivity through five factors: interaction engagement, respect of cultural differences, interaction confidence, interaction enjoyment, and interaction attentiveness |
| Scale of Ethnocultural Empathy (SEE) | Wang et al. (**2003**) | Self-report; 6-point Likert scale (*strongly disagree that it describes me* to *strongly agree that it describes me*) | Online survey | 31 items | Measures empathy toward people of racial and ethnic backgrounds different from one's own. Contains four subscales: empathic feeling and expression, empathic perspective taking, acceptance of cultural differences, and empathic awareness. |
| Multicultural Personality Question | Van der Zee and Van Oudenhoven (**2000**) | Self-report; 5-point Likert scale (*not at all applicable* to | Online survey | 78 items | Measures multicultural effectiveness through five subscales: cultural empathy, |

| | | | | | |
|---|---|---|---|---|---|
| naire (MPQ) | | *totally applicable*) | | | open-mindedness, emotional stability, flexibility and social initiative. |
| Beliefs, Events, and Values Inventory (BEVI) | Shealy (2004) | Self-report and biographical data | Online survey | ؟ | Measures openness to transformational experiences such as international educational experiences through 10 process scales, such as negative life events and need for control |
| Cultural Orientations Indicator (COI) | Schmitz, Tarter, and Sine (2012) | Self-report; response scale unknown | Online survey | ؟ | Assesses cultural preferences across three dimensions: interaction style, thinking style, and sense of self. Provides the test taker with comparisons of their own scores to country norms as well as recommendations for further learning and |
| Culture in the Workplace Questionnaire | Hofstede (2010) | Self-report | Online survey | 60 items | Based on Hofstede's five cultural dimensions: individualism, power distance, certainty, achievement, and time orientation. Designed to serve as a cultural |

| | | | | | |
|---|---|---|---|---|---|
| | | | | | values-based self-awareness |
| Global Awareness Profile | Corbitt (**1998**) | Performance measure (knowledge test) | Online survey | 126 items | Includes two dimensions: geography and context. Subcategories of context include environment, politics, geography, religion, socioeconomics, |
| Global Perspectives Inventory (GPI) | Global Perspective Institute (GPI) | Self-report; 5–point Likert scale; *strongly agree* to *strongly disagree* | Online survey | 3 forms (general student, new student, study abroad posttest); 35 items; 6 subscales with 4–7 items per scale | Measures how college students relates to others from backgrounds different from their own and how they perceive their own cultural heritage. Measured through three dimensions and six global perspective scales: cognitive (with knowing and knowledge scales), intrapersonal (with identity and affect scales), and interpersonal (with social responsibility and social interactions scales |
| Intercultural Compete | Elmer (**1987**) | Self-report; response | Online survey | 45 items | Measures intercultural effectiveness |

| | | | | | |
|---|---|---|---|---|---|
| ncy Scale (ICS) | | scale unknown | | | through 12 factors, such as approachable, intercultural receptivity, positive orientation, forthrightness, social openness, enterprise, shows respect, flexibility, perseverance, cultural perspectivism, venturesome, and social confidence |
| Tests for Hidden Bias | Project Implicit https://implicit.harvard.edu/implicit/takeatest.html | Performance measure (implicit association tests) | Online survey | 14 different tests | Implicit association tests that measure unconscious biases such as negative prejudices toward various ethnic |
| Miville-Guzman Universality–Diversity Scale (M–GUDS) | Fuertes (**2000**) | Self-report; 6-point Likert Scale; *strongly disagree* to *strongly agree* | Online Survey | 45 questions in the long form; 15 questions in the short form | Measures universal–diverse orientation (UDO), or the degree to which a person accepts diversity among people, through three subscales: diversity contact, relativism appreciation, and comfort with difference. |
| Cross-Cultural World-Mindedn | Der-Karabetian (**1992**) | Self-report; response scale unknown | ؟ | .؟ | Measures worldmindedness |

| | | | | | |
|---|---|---|---|---|---|
| ess Scale (CCWM) | | | | | |
| Multicultural Awareness–Knowledge Skills Survey (MAKSS) | D'Andrea, Daniels, and Heck (**1991**) | Self-report; 4-point Likert scale; *strongly disagree* to *strongly agree* | Paper and pencil | 60 items | Designed for multicultural counseling; measures an individual's multicultural awareness, knowledge, and skills. |
| Behavioral Assessment Scale for Intercultural Effectiveness (BASIC) | Koester and Olebe (**1989**) | Peer rating; 4-point rating scale | Peer rating; 4-point rating scale | 9 items | Measures intercultural communication effectiveness through peer ratings |
| Global Team Process Questionnaire (GTPQ) | Bing (**2001**) | Self-report: Likert items as well as narrative questions | Paper and pencil | ؟ | Measures effectiveness in global teams by examining skills, attitudes, and processes. |
| Inventory of Cross-Cultural Sensitivity (ICCS) | Cushner (**1986**) | Self-report: 7-point Likert scale; *strongly disagree* to *strongly agree* | Paper and pen/online survey | 32 items (5 subscales, 5–10 items per subscale) | Measures cultural integration, behavioral response, intellectual integration, attitudes toward others, and empathy |

| | | | | | |
|---|---|---|---|---|---|
| Implicit Association Test | Bazgan and Norel (**2013**) | Performance measure (implicit association tests) | Online test | ؟ | Implicit measure of ICC with categories of national or minority language. Categorized stimuli were represented by the names of multiethnic localities from Romania, presented in the national language, Romanian; and minority languages: Hungarian, German, Turkish, Greek, and Slavonic. |
| Global Competence Aptitude Assessment | W. D. Hunter et al. (**2006**) | Performance measure (multiple-choice) | Online test | ؟ | Measures internal readiness (self-awareness, willingness to take risks, open-mindedness, and perceptiveness/respectfulness of diversity) and external readiness (global awareness, world history knowledge, intercultural competence, and effectiveness across cultures). |
| Cross-Cultural Sensitivit | Pruegger and Rogers (**1993**) | Self-report: 6-point Likert scale; *strongl* | Paper and pencil | 24 items total (two equivalen | Measures the valuation and |

| | | | | | |
|---|---|---|---|---|---|
| y Scale (CCSS) | | *y disagree* to *st rongly agree* | | t forms with 12 items each) | tolerance of different cultures. |
| Intercultu ral Commun ication Compete nce (ICC) | Arasaratnam and Doerfel (**2005**) and Arasaratnam (**2009**) | Self-report: 7-point Likert scale; *strongl y disagree* to *st rongly agree* | Paper and pencil | 10 items; 3–4 items for each dimensio n | Cognitive, affective, and behavioral dimensions of intercultural communication competence |
| Intercultu ral Sensitivit y Inventory (ICSI) | Bhawuk and Brislin (**1992**) | Self-report: 7-point Likert scale; *very strongly disagree* to *ve ry strongly agree* | Paper and pencil | 46 items; 14–16 items per subscale; individua lism versus collectivi sm are asked in relation to own or other culture | Measures individualism versus collectivism and flexibility/open-mindedness |
| Global Compete ncies Inventory | Kozai Group; Bird et al. (**2002**) Stevens, Bird, Mendenhall, and Oddou (2014) | Self-report: 5-point Likert scale; *strongl y disagree* to *st rongly agree* | Online test | 160 items; 16 subscales with items ranging from 6–14 | Competencies can be loosely grouped into perception, relationship, and self-management. |
| Cross-Cultural Social Intelligen ce | Ascalon et al. (**2008**) | SJT: 4 response options | ؟ | 14 scenarios ; replies vary across ethnocent ric–nonethno centric | Measures knowledge, skills, and other characteristics that promote successful social interaction in cross-cultural interactions. |

| | | | | | |
|---|---|---|---|---|---|
| | | | | and empatheti c– nonempat hetic | |
| Cultural Intelligen ce Assessme nt | Thomas et al. (2015) | Self-report (multiple response scales) and verbal protocol trace | ؟ | 24 items plus verbal trace protocol | Measures cultural knowledge, knowledge complexity, cultural metacognition (self-report and trace), relational skills, perceptual acuity, empathy, adaptability, and tolerance for uncertainty. |
| Nonverba l Commun ication Compete nce Scale (NVCCS ) | Kupka and Everett (2008) | Self-report; anchors unknown | Paper and pencil | 5 items | behaviors of foreign culture members, the skills to show nonverbal behaviors, and the motivation to interpret and present them. Additionally, appropriateness and effectiveness in nonverbal communication is |

## 3. Reliability and Validity Evidence of Available Assessments

Based on the standards for Educational and Psychological Tests (American Educational Research Association {AERA}, American Psychological Association {APA}, & National Council on Measurement in Education {NCME}, 2014), each instrument should: a) create accurate and consistent scores (reliability) and, b) produce adequate proof to support that the test measures what it claims to measure (validity). In this section, first, I discuss reliability related to ICC assessment instruments reviewed in this study. then, the validity evidence regarding the internal structure, the relationship with conceptually related constructs, and the

relationship with criteria is discussed. In table 3, a summary of the validity and reliability is presented.

### 3.1 Test and Scale Reliability

As aforementioned, most ICC assessments include particularly Likert type items with adequate reliability evidence. More than 90% of the scales presented are assessed with coefficient alpha (α) which indicates an evidence of adequate reliability. Nevertheless, to assess ICC with more than on sub domain, a number of assessments such as Cross-Cultural Adaptability Inventory {CCAI} (Davis & Finney, 2006) with sufficient overall alpha values consisted of subscale scores that fall below 70, which is an acceptable cut off (Kline, 2000). However, a few numbers of scales could indicate an adequate reliability using test and retest (e.g., Inventory of Cross-Cultural Sensitivity; Baz gan & Norel, 2013). See Table 2 for scale specific reliability information.

### 3.2 Validity Evidence Regarding Internal Structure

The internal structure of the assessments (i.e., dimensional) is a predominant aspect of validity evidence. The internal structure represents if the connection between test items harmonize with one or more intended domains (or dimensions of the assessment (AERA, et al., 2014). Confirmatory Factor Analysis is one of the most commonly used methods to measure the internal structure (CFA; Rios & Wells, 2014). More than 10 assessments presented in table 2, have a single overall score to examinees, and five of these assessments give evidence to support the dimensional structure of the assessment. The one-factor model was fitted to data from the Nonverbal Communication Competence Scale and the outcome suggested that all items were measuring the same construct (Grafand Mertosacker, 2009). The same evidence for the Intercultural Communication competence Test and the basic test was provided by Arasaratnam (2009) and Olebe and Koester (1989).

About half of the assessments with subscale scores provide evidence to support the multidimensional structure of the assessments. The S.F A outcomes, for example, indicated the intended constructs were measured by the four subscales of the scale of Ethnocultural Empathy, and the four factors shared nearly 81% of the total variance (Wang et.al 2003). An appropriate model of a five-factor for Intercultural Development Inventory was suggested by Hammer et.al (2003). However, data may not be able to support a multidimensional structure of assessments all the time. For example, a loose support for the four-factor model which has suggested for the CCAI was found by Davis and Finney (2006). Nguyen, Biderman, and McNary (2010) also found that each item from the CCAI loaded on a general factor namely Cross-Cultural Adaptability and one of the nine group factors (e.g., personal autonomy, resilience, flexibility/openness, and the like). These group factors provided evidence for constructs which were not explained by the general factor. Thus, although the CCAI included four subscale scores, the outcomes from the two studies did not support a four-dimensional structure of the assessment. As a result, it is clear that the evidence for multidimensional structure for the present ICC measures is not strong enough. In addition, about half of the ICC assessments introduced in this paper did not present sufficient internal structure. Providing this evidence is supported by best practices for scale construction by representing an appropriate model fit of an item-level factor analysis. For instance, the Global Competencies Inventory (GCT; Bird, Stevens, Mendenhell & Oddou, 2002) presented only the correlation among the three sub scores

in place of the internal structure of assessment. An important gap in the validity evidence and thus an especially considerable weakness is demonstrated by the lack of evidence explaining the structure of the scale.

### 3.3 Validity Evidence Concerning Relationships with Conceptually Related Constructs

The relationship with conceptually related constructs is the second aspect of validity evidence which is traditionally called convergent and discriminate validity. A correlation coefficient between two assessments is commonly used to measure the degree to which the constructs assessed by the two assessments are related to each other. A valid measure would indicate harmonious with related constructs and difference with irrelevant constructs. Since the correlation coefficient is influenced by the reliability of the two assessments (i.e., low reliability would lower the correlation coefficient below the level it would have reached when the reliability is high), it is vital to mention the reliability information beside the correlation coefficient. Totally, about half of the present ICC measures presented in this paper introduced some evidence regarding a relationship with related constructs.

Typically, the Cultural Intelligence Construct includes pretty well evidence from organizational samples in educational contexts (Leung et al., 2014). Erez et al., (2013), Lizak & Erez, (2015) carried out two studies exerting the Cultural Intelligence scale (Ang, Van Dyne, & Koh, 2006; Ang et al; 2007) with students taking part in a cross-cultural virtual team project. The outcomes indicated a strong relationship (r□=□.50) between the cultural intelligence of students in global virtual context namely global identities (Erez & Gati, 2004).

The researchers assessed global identities with a validated and adequately reliable global identity scale (α□=□.85; Erez & Gati, 2004; Shokef & Erez, 2006, 2008). One of the studies further related cultural diversity (r□=□.16) and leadership emergence (r□=□.56; Lisak & Erez, 2015). Hammer et al., (2003) conducted a study and they concluded that there are relationships among the subscales of International Development Inventory (IDI; α□=, 80-,85) and two related assessments- the world- mindedness scale (α□=□. 86). Higher scores on the denial/defence subscale of the IDI were relevant to lower scores on the world-mindedness scale (r□=□,29) and higher scores on the Intercultural Anxiety Scale (r□=□, 16).

Structural equation modeling separates the latent construct and organizes another robust method of supporting relationships among assessments. Nguyen et al (2010) applied a structural equation modeling technique to capture the relationship between the CCAI and Golberg's (1999) IPIP Big five questionnaire instead of calculating the correlation coefficient from observed scores. The outcomes proved to be weak to moderate correlations between the two assessments (r□=□, 18-, 55), showing that students with an appropriate form of Cross-Cultural Adaptability are more likely to be agreeable, emotionally stable, extroverted open to new experiences, and conscientious. The correlation coefficient measured from the structural equation model is the correlation between the underlying constructs of two assessments measurement error, unlike the statistics used in the Hammer et al (2003) study, does not influence the structural equation model correlation. Hence, it has been proved that a promising method for future research to allow for validity information concern g relationship with conceptually related constructs is structural equation modeling.

### 3.4 Validity Evidence Concerning Relationship with Criteria

Another important aspect of validity evidence is the relationship between the related s and the assessment (AERA et al., 2014). Self-evaluation, job performance, and peer impressions and the like are examples of the criteria applied for existing ICC measures. This type of validity evidence has been provided by few assessments in table 2, because of the resources-heavy requirements of criterion data collection.

Nguyn et al., (2010) researched if the subscale score of the CCAI would prognosticate the number of international job assignments while controlling for the variance of the general factor (Cross-Cultural Adaptability). The outcomes supported the hypothesis to some extent as only two subscales (resilience and personal autonomy) was weakly correlated with the logarithm number of international job assignments ((r□=□, 20 and r□=□, 20, respectively), and no subscale were correlated with the actual number of assignments. Matsumoto et al. (2001) examined the participants who took the Intercultural Adjustment Potential Scale (ICAPS) also rated themselves and all their members of the focus group on a two-item rating scale about Intercultural Adjustment. Both ratings of all participants were made by two interviewers. The analysis indicated that the composite score of the ICAPS was considerably correlated with self, peer, and interviewer ratings ((r□=□, 90, .70, and .66, respectively; P≤.001), supporting the usefulness of the ICAPs in explaining Intercultural Adjustment. Furthermore, the Miville-Guzman University-Diversity scale, which assesses awareness and potential acceptance of both similarities and differences in others, was not considerably relevant to the SAT verbal scores (Miville et al., 1999), which allow for evidence of discriminant construct validity construct validity. Nevertheless, in a study by UK students in culturally different terms, the multicultural personality questionnaire was found to be relevant to exam grades (Vander Zee, Atsma, & Broabeck, 2004).

Hammer (2011) conducted a study with 71 recruiters in a U.S. high-tech organization, scores on the IDI were found to be correlated (r□=□, 43) with the rating of success in meeting variety purposes for recruitment. In another funded study on study abroad students, 1.500 students completing a 10-months homestay program developed by AFS intercultural programs, an American based study abroad facilitator, were compared to a control group (r□=□, 638) of students who remained at their home institutions. Students who participated in the homestay program inhabited in Brazil, Ecuador, Hongkong; Japan, Italy, Costa Rica, United States, and Austria. It was found that the scores were positively correlated with the number of intercultural friends students provide evidence of having a sociometric measure of experience success showing the ability of students to make international relational networks (Hammer, 2005). The assessment provides evidence that anxiety reduces but satisfaction increases with the experience.

It is suggested by other evidence the cultural intelligence scale (CQS) may relate to a number of valued student results. Higher score on the CQS were particularly related to commitment to and satisfaction with international educational courses (e.g., Morell, Ravlia, Ramsey, & Ward, 2013; Ramsey, Barakat, & Aad, 2014), global virtual team leadership (Erez et al., 2013; Lisak & Erez, 2015), intention to work abroad (e.g., Remhof, Gunkel, & Schlaege, 2013). These results which are part of the previous experience category, are valuable criteria because they have been related to global leadership effectiveness (e.g., Caligiuri & Tarique, 2012).

It has been researched that study abroad experiences increase student competencies by exploiting this scale. (Engle & Crowne, 2014; Varela & Gatlin-Watts, 2013). Nevertheless, the validity evidence relating to scale mixes with adjustment while studying abroad. In a study conducted with international students studying in New- Zealand, the result indicated that the

motivational subscale was not predictive of psychological adjustment during study abroad (Ward, Wilson, & Fischer, 20011); in another study conducted by Lin, Chen, Song, (2012), the outcome showed that cultural intelligence was not related to adjustment. It is clear that the two studies applied various subscales for adjustment-the Socio-cultural Adaptation Scale (Ward & Kennedy, 1999) and the Black and Stephans (1989) scale assessing general adjustment, interactional, and work. Although the Black and Stephans Scale has several measurement concerns such as proper validation evidence, but it is commonly used by researchers (Thomas & Lazarova, 2006).

**Table 2.** Reliability and Validity Evidence of Intercultural Competence (ICC) Assessments

| Test | Reliability | Validity |
|---|---|---|
| Cross-Cultural Adaptability Inventory (CCAI) | Across all four subscales, alpha = .68–.90; alpha of .90 for entire scale. | Internal structure: EFA failed to identify an interpretable structure and CFA found poor fit of four-factor structure (Davis & Finney, **2006**). In another study, both the one-factor model and the four-factor model fit the data poorly, and four subscales were highly correlated with each other after controlling for common method variance, suggesting lack of differentiation among the subscales (Nguyen et al., **2010**). Relationship with other assessments: The four subscales of the CCAI have low to moderate correlation with Goldberg's IPIP Big Five questionnaire ($r$ = .182 to .548, $p$ < 0.05) from Nguyen et al. (**2010**). Relationship with criteria: Emotional resilience subscale and personal autonomy subscale can weakly predict the number of international job |

| | | |
|---|---|---|
| | | assignments (Nguyen et al., **2010**). |
| The Global Perspective Survey | Test–retest ranging from *r* = .49 (cognitive–knowledge subscale after 3 weeks) to .81 & interpersonal–social<br><br>responsibility; alphas ranging from .657 (cognitive–knowing) to .773 (cognitive–knowledge) | Relationship with other assessments: *t* tests revealed that initial scores on the Global<br><br>Perspective Survey were significantly different (at the 0.05 level, except the perceptual acuity subscale) from the CCAI (Smith and Mitry, **2008**). |
| Assessment of Intercultural Competence (AIC) | Overall: alpha = .824. Subscale: alpha = .86–.98 (Fantini & Tirmizi, **2006**); subscale: alpha = .59–.73 (Almeida, Simões, & Costa, **2012**). | Internal structure: For the first component, *knowledge*, principal component analysis suggested two underlying factors. Consequently, the items were collapsed into two clusters according to factor loadings. In each of the remaining three components (*attitude*, *skills*, and *awareness*), however, most items loaded onto aSingle factor. In a few cases, where it was found that items loaded onto two factors at the same time, these items were excluded. Their exclusion led to single component loadings and showed an improvement in the explained variance |
| Intercultural Adjustment Potential Scale (ICAPS) | Alpha = .78. | Relationship with other assessments: ICAPS was significantly correlated to three scales of the SAS, four scales of the SCBAI, the Beck Depression Inventory, and the |

| | | |
|---|---|---|
| | | Adjustment Scale. ICAPS also demonstrated significant correlations with the CCAI, the Big Five Inventory, and the Million Clinical Multiaxial Inventory (MMCI). Relationship with criteria: Composite scale score was found to be significantly correlated with self-rating, peer rating, and facilitator rating of adjustment. |
| Cultural Intelligence Scale (CQS) | Reliabilities exceeded 0.70 (metacognitive CQ = 0.77, cognitive CQ = 0.84, motivational CQ = 0.77, and behavioral CQ = 0.84). | Internal structure: Used CFA to confirm four-dimensional structural of the 20 items. In cross-cultural reliability studies, CFA maintained acceptable fit across samples. Relationship with other assessments: Eleven of 16 correlations between CCAI and CQS were significant. Discriminant validity demonstrated with CCAI ($r$ = .07 to .48, mean = .22), FFM: Five-Factor Model of Personality ($r$ = −.08 to .28), EI: emotional intelligence (USA: $r$ = .18 to .41, mean = .27; Singapore: $r$ = .12 to .28, mean = .18), and CJDM: cultural judgment and decision making ($r$ = .13 to .27). |
| Global Competencies Inventory (GCI) | Subscale alpha = .72–.92. | Relation with other variables: Correlated (across three subscales, $r$ = 0.12–0.29) with the Worldmindedness Scale (Sampson & Smith, **1957**; |

| | | |
|---|---|---|
| | | Wiseman, Hammer, & Nishida, **1989**). Also correlated (across three subscales, *r* = 0.13 to<br><br>0.16) with the Intercultural Anxiety Scale (Stephan & Stephan, **1985**).<br>15/16 factors correlated with neuroticism (*r* =.20 –.69); 8/16 with extraversion (*r* =.21–42); 16/16 with openness (*r* =.20 –.64); 13/16 with agreeableness (*r*=.15 –.46), and 9/16 with conscientiousness (*r* =.13 –.44). |
| Intercultural Development Inventory (IDI) | The reliability results are denial/defense scale (14 items, alpha = .85), reversal scale (nine items, alpha = .80), minimization scale (10 items, alpha = .85), acceptance/adaptation scale (14 items, alpha = .84), and encapsulated marginality scale (five items, alpha = .80). | Internal structure: Confirmatory factor analysis narrowed items to 52, distributed across five factors: denial/defense, reversal, minimization, acceptance/adaptation, and encapsulated marginality. Relationship with other assessments: IDI scales significantly correlated with Worldmindedness Scale Validity (DD *r* = −.29, AA *r* = .29, CM *r* = .12) and Intercultural Anxiety Scale (DD *r* = .16, AA *r* = −.13, EM *r* = .14). Assessment fairness: No significant differences on IDI for gender, age, education, or social desirability |
| Intercultural Sensitivity Scale (ISS) | Cronbach's alpha for scale = .86. | Internal Structure: Five factors had eigenvalues higher than 1, accounting |

| | | |
|---|---|---|
| | | for 37.3% of the variance. Relationship with other assessments: ISS is correlated with Interaction Attentiveness Scale $r = .20$, Impression Rewarding Scale $r = .41$, Self-Esteem Scale $r = .17$, Self-Monitoring Scale $r = .29$, Perspective Taking Scale $r = .52$, Intercultural Effectiveness Scale $r = .57$, and Intercultural Communication Attitude Scale $r = .74$ (all with $p$ values, $< 0.05$). |
| Scale of Ethnocultural Empathy (SEE) | Alphas of .91, .89, .75, .73, and .76 were obtained for the SEE total, EFE, EP, AC, and EA. | Internal structure: The four factors were well constructed, and the four factors shared approximately 81% of the total variance. Relationship with other assessments: highly correlated with the M–GUDS, or Miriville–Guzman Universality–Diversity Scale ($r = .70$, $p < 0.05$); the IRI, or Davis Interpersonal Reactivity Index ($r = .42$ to .48, $p < 0.05$); and the BIDR, or Balanced Inventory of Desirable Responding ($r = .23$, $p < 0.05$). |
| Multicultural Personality Questionnaire (MPQ) | Subscale alpha = .68–.87 | Internal structure: Four factors with eigenvalues greater than 4 emerged. Relationship with other assessments: Correlations with Big Five and Need for Change were significant at $p < 0.05$ except flexibility with |

| | | |
|---|---|---|
| | | agreeableness and conscientiousness; emotional stability with openness to<br><br>experience; emotional stability with need for change, and rigidity only significantly correlated (negatively) with flexibility. |
| Beliefs, Events, and Values Inventory (BEVI) | Subscale alpha = .62–.95 | Internal structure: EFA clustered 494 items into 10 *process scales*. Relationship with criteria: Evidence of validity is indicated by a number of studies demonstrating that the BEVI is able to predict group membership across a wide range of demographic variables, including gender, ethnic background, parental income, and political orientation (cf. Hayes, Shealy, Sivo, & Weinstein, **1999**; Isley, Shealy, Crandall, Sivo, & Reifsteck, **1999**; Shealy, Burdell, Sivo, Davino, & Hayes, **1999**; Shealy, Sears, Sivo, Alessandria, & Isley, **1999**). |
| Cultural Orientations Indicator (COI) | No reliability information available. | Internal structure: Factor analysis revealed that COI scales map onto three or four distinct dimensions: interaction style, thinking style, and sense of self. Continua are aligned with these dimensions. |
| Culture in the Workplace Questionnaire | Hofstede (2010) | Relationship with criteria: Cultural values are just as |

| | | |
|---|---|---|
| | | robust as personality traits and demographics in predicting individual outcomes (e.g., organizational commitment, identification, and citizenship behaviors). |
| Global Awareness Profile | Test–retest reliability for 56 undergraduate students was 0.83, $p < 0.01$. | Face validity achieved through consultation with regional and subject experts at the university level. No predictive or comparative validity evidence sought. Discriminant construct validity demonstrated through ANOVA with 71 test takers, some with no cross-cultural experience, others with some; those with at least one month's experience scored significantly higher (80 vs. 66 correct answers). |
| Global Perspectives Inventory (GPI) | Subscale alpha = .66 –.77. | Internal structure: Principal component analysis using varimax rotation revealed six factors with eigenvalues higher than 1, accounting for 50% of cumulative variance. Relationship with other assessments: Research conducted by Anderson and Lawton (**2011**) concluded that IDI and GPI do not measure similar characteristics. |
| Intercultural Competency Scale (ICS) | Strubler, Agarwal, Park, and Elmer (**2011**) | Relationship with other assessments: Correlations between ICS and CCSI: |

| | | |
|---|---|---|
| | | approachable ($r = .30$), perseverance ($r = .34$), cultural perspectivism ($r = .40$), venturesome ($r = .35$); all were at least $p < 0.05$ level. |
| Tests for hidden bias | Not significantly different from other IATs | Relationship with other assessments: not significantly different from other IATs |
| Miville-Guzman Universality–Diversity Scale (M–GUDS) | Alphas range between .89 and .94. | Internal structure: Analysis yielded a factor structure composed of a large general factor along with two smaller factors. Patterns of correlations of the factor analyzed M–GUDS with several other measures closely mirrored those of the original scale. These findings generally supported a unidimensional structure of the M–GUDS. They also indicated that the total scale score, rather than subscale scores, should be used to reflect the instrument's apparent unidimensional nature. Relationship with other assessments: M–GUDS significantly associated with White Racial Identity Attitude Scale (WRIAS): autonomy ($r = .48$), contact ($r = .45$), disintegration ($r = -.56$), reintegration ($r = -.60$), and pseudo-independence ($r = .42$). M–GUDS also significantly negatively correlated with dogmatism scale (−.27) and homophobia scale (−.33), $p < 0.01$. Relationship with criteria: |

| | | |
|---|---|---|
| | | not correlated with SAT (Miville et al., **1999**). |
| Cross-Cultural World-Mindedness Scale (CCWM) | Cronbach's alphas range among 10 countries' samples between .69 and .88. | Relationship with criteria: Subsequent analysis suggested criterion validity for political party orientation. |
| Multicultural Awareness–Knowledge Skills Survey (MAKSS) | Reliability for subscales: awareness (alpha = .75), knowledge (alpha = .90), skills (alpha = .96). | Internal structure: Factor analysis suggested that awareness might have a three-factor solution, but knowledge and skills were both satisfied with a one-factor solution. Intercorrelations: awareness and knowledge $r = .45$; awareness and skills $r = .32$; knowledge and skills $r = .51$. |
| BASIC | Reliability for whole scale alpha = .80. | Internal structure: Factor analysis revealed one underlying factor solution with an eigenvalue of 3.85. Relationship with other assessments: correlation with global measure of effectiveness, $r = .60$. |
| Global Team Process Questionnaire (GTPQ) | No reliability information available | Relationship with criteria: The assessments results mirrored findings from interviews. |
| Inventory of Cross-Cultural Sensitivity (ICCS) | Overall: Alpha = .85 for Canadian sample and .77 for Japanese sample.<br><br>Subscale: alpha = .37–.73 for Canadian sample and .25–.55 for Japanese sample. | Internal structure: a moderate fit five-factor solution from both the<br><br>Canadian data and the Japanese data |

| | | |
|---|---|---|
| Implicit Association Test | Test–retest ($n = 71$) $r = .77$. | Relationship with criteria: Weighted average of IAT-criterion correlations (ICCs), based on 122 reports that contained 184 independent samples, was *rICC* = .274. For socially sensitive topics, the predictive validity of self-report measures was remarkably low and the incremental validity of IAT measures was relatively high. |
| Global Competence Aptitude Assessment | No reliability information available. | Surveyed international educators as well as human resource professionals at multinational corporations to identify critical elements of global competence. General agreement between groups, with some exceptions |
| Cross-Cultural Sensitivity Scale (CCSS) | Internal consistency alpha: .93. In subsequent studies, two parallel forms developed, with alphas of .87 and .80. | CCSS scores correlated with verbal IQ and full scale IQ among students in Grades 3, 5, and 6 (Klein, **1995**). |
| ICC | Cronbach's alpha = .77, $M = 4.79$, $SD = .88$. | Internal structure: one-factor solution Relationship with other assessments: Correlation analysis revealed positive relationships between ICC and attitude toward other cultures [$r(302) = .51$, $p = .01$], ICC and motivation [$r(302) = .50$, $p = .01$], and ICC and interaction |

| | | |
|---|---|---|
| | | involvement [$r(302) = .54$, $p = .01$], and a negative correlation between ICC and ethnocentrism [($r(302) = -.62$, $p = .01$]. |
| Intercultural Sensitivity Inventory (ICSI) | Alpha for College of Business sample: .82;<br><br>Alpha for East–West Center sample: .84. | Internal structure: There are two factors: collectivism and individualism. |
| Nonverbal Communication Competence Scale (NVCCS) | Coefficient alpha = .87. | Internal structure: one-factor solution with high loading items Relationship with other assessments: Self-assessment demonstrated significant positive correlations with nonverbal communication competence ($r = .514$), and praising of others and ability to deal with compliments ($r = .398$), intercultural sensitivity ($r = .263$), openness/flexibility ($r = .308$), display of negative feelings |
| Cultural Intelligence Assessment (CIA) | Coefficient alpha =.68 for full measure, .61 for empathy subscale, .71 for ethnocentrism | Relationship with other assessments: Expected positive significant correlation with preexisting empathy and ethnocentrism scales for overall and subscales (average $r$ =.20). Also related to conscientiousness, emotional stability, and openness (average $r$ −.30); not related to tolerance for ambiguity or self-monitoring. |
| | | |

| | | |
|---|---|---|
| Cross-Cultural Social Intelligence (CCSI) | Coefficient alpha = .95 for cultural knowledge, = .90 for knowledge complexity, = .82 for self-report metacognition, = .79 for verbal protocol trace, = .73 for relational skills, = .69 for perceptual acuity, = .66 for empathy, = .70 for adaptability, = .56 for tolerance of uncertainty. | Relation with criteria: All factors except for adaptability positively related to intercultural effectiveness (unvalidated composite of task completion in intercultural settings, development of good interpersonal relations, and feelings of well-being while interacting with culturally different others). |

## 4. Discussion & Conclusion

It has been identified that ICC as one of the most vital life skill is likely to forecast favorable outcomes in the 21st-century workforce. Because higher education institutions start to explore the traditional models of learning results and put emphasis on these life skills, it is a critical need to measure if students have these important competencies. Furthermore, assessments are necessary to specify if the students' abilities and skills regarding ICC progress during their studies.

Surveys and portfolio assessments were two important assessment formats introduced in this study; however, the portfolio may not be an appropriate assessment format because it fails to standardize students' work products and also ensure inter-rater reliability in students' works scoring. Thus, a survey is considerably more standardized and norm-referenced to meet students' needs. As it was shown different assessment formats such as Likert-Scale items, Multiple-Choice Items, Implicit Association Tests and Q-Sort Methodology, Situational Judgment Tests, and Simulation-Based measurement were introduced in this study. As it was shown each assessment format has its own weakness and strength. Most ICC assessments in Table 1 attempted to capture components of self-report Liker items.

As it was clear, all the assessments in Table 2, included satisfactory reliability at the test level; nevertheless, there are still few deficits. It was found that subscale score reliability of five assessments was unsatisfactory ($\alpha \leq .70$), such as the cultural Intelligence Assessment, CCAI, and the Global Perspective Inventory. When subscale scores are used for diagnostic purposes, unreliable sub scores produce inaccurate diagnoses and false information for individuals. Unreliable subscales mean that error will infect various aspects unevenly and decrease the quality of a development plan made according to scores. In addition, when some subscale scores randomly fluctuate, it would be not easy to validate ICC training interventions. Another issue is related to the comparability among test forms. In table 2, there are three assessments that involve more than one test form, two included high correlation between test forms, although one did not provide any information. The quantity and quality of validity evidence, unlike the reliability evidence, changed considerably among present ICC measures. In table 3, nearly half of the assessments provided evidence for validity concerning internal structure, less than one

third represented evidence concerning the relationship with related criteria, and only two assessments introduced all three aspects of validity evidence.

Not only the quantity but also the quality of some present validity evidence was also unacceptable. For example, the hypothesized data did not support the hypothesized internal structure of some assessments, raising the question about subscale score reporting. As a result of the low reliability of the tests, the relation between some ICC assessments and their related measures were not measured strongly.

Totally, some assessments that developed after 2000 (e.g., the Cultural Intelligence Scale and the IDI) and the assessment designed by organizations (e.g., the CCAI) had stronger validity evidence. However, the assessments designed by independent researchers reported inadequate validity evidence. The problem with insufficient validity evidence is related to the financial problems or less available statistical packages, and obsolete approach to validity. After Messic (1995) definition of validity as a single construct for which researchers could provide different kinds of evidence, developers acknowledged the significance of collecting a range of validity evidence to support test score inferences. In recent years, more validity research has been conducted; however, one aspect of validity that is still missing is its relationship with criteria. According to Messic, not any type of evidence is important, although, lack of criteria-related evidence should not be ignored. Very few assessments were related to any sort of accepted criteria. Thus, future validity research should be supported to collect criteria information to explain the extent to which the scores from an ICC assessment prognosticate respondents' skills to communicate and work across cultures in real situations. Criterion related validity is compelling in terms of investment. The persuasive evidence of their relations to valued results will be the best foundation if the strong argument is to be created for higher education to invest in the development of these skills.

## References

Almeida, J., Simões, A. R., & Costa, N. (2012). Bridging the gap between conceptualisation & assessment of intercultural competence. *Procedia-Social and Behavioral Sciences, 69,* 695-704.

American Educational Research Association, American Psychological Association, & National Council On Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington,DC: Author.

Ang, S., Van Dyne, L., & Koh, C. (2006). Personality correlates of the four-factor model of cultural intelligence. *Group & Organization Management, 31*(1), 100–123.

Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation and task performance. *Management and Organization Review, 3*(3), 335–371.

Arasaratnam, L. A. (2008). Acculturation process of Sri Lankan Tamil immigrants in Sydney: An ethnographic analysis using the Bidirectional model (BDM). *Australian Journal of Communication, 35*, 57–68.

Arasaratnam, L. A. (2009).The development of a new instrument of intercultural communication competence. *Journal of Intercultural Communication, 20*, 1–11

Ascalon, M. E., Schleicher, D. J., & Born, M. P. (2008). Cross-cultural social intelligence: An assessment for employees working in cross-national contexts. *Cross Cultural Management: An International Journal, 15*(2), 109–130.

Bazgan,M., & Norel,M. (2013). Explicit and implicit assessment of intercultural competence. *Procedia- Social and Behavioral Sciences, 76*(15), 95–99. doi:10.1016/j.sbspro.2013.04.080

Bhawuk, D. P. (2001). Evolution of culture assimilators: Toward theory-based assimilators. *International Journal of Intercultural Relations, 25*(2), 141–163.

Bhawuk, D. P., & Brislin, R. (1992).The measurement of intercultural sensitivity using the concepts of individualism and collectivism. *International Journal of Intercultural Relations, 16*, 413–436.

Bhawuk, D. P., & Brislin, R. (2000). Cross-cultural training: A review. *Applied Psychology, 49*(1), 162–191.

Bing, J. W. (2001, February). *Developing a consulting tool to measure process change on global teams: The Global Team ProcessQuestionnaire™*. Paper presented at the national conference of the Academy of Human Resource Development, Tulsa, OK.

Bird, A., Stevens, M. J., Mendenhall, M. E., & Oddou, G. (2002). *The global competencies inventory*. The Kozai Group, Inc., St. Louis, MO.

Caligiuri, P., & Tarique, I. (2012). Dynamic cross-cultural competencies and global leadership effectiveness. *Journal ofWorld Business, 47*, 612–622.

Corbitt, J. N. (1998). Global awareness profile (1). Yarmouth, ME: Intercultural Press.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, *82*(1), 143–159.

Chen, G. M. (1992). A test of intercultural communication competence. *Intercultural Communication Studies, 2*, 62–83.

Chen,G.M.,&Starosta,W. J. (1996). Intercultural communication competence:Asynthesis. *Communication Yearbook*, *19*(1), 353–384.

Chen, G.M., & Starosta,W. J. (2000).The development and validation of the Intercultural Sensitivity Scale. *Human Communication, 3*, 1–15.

Corbitt, J. N. (1998). *Global awareness profile* (*1*). Yarmouth, ME: Intercultural Press.

Cushner, K. (1986).The Inventory of Cross-Cultural Sensitivity. Kent,OH: School of Education, Kent StateUniversity.

Davis, S. L., & Finney, S. J. (2006). A factor analytic study of the Cross-Cultural Adaptability Inventory. *Educational and Psychological Measurement, 66*(2), 318–330.

D'Andrea,M., Daniels, J., & Heck, R. (1991). Evaluating the impact ofmulticultural counseling training. *Journal of Counseling&Development, 70*(1), 143–150.

Deardorff, D. K. (2006). Identification and assessment of intercultural competence as a student outcome of internalization. *Journal of Studies in International Education, 10*(3), 241–266. doi:10.1177/1028315306287002

Der-Karabetian, A. (1992).World-mindedness and the nuclear threat: A multinational study. *Journal of Social Behavior & Personality, 7*(2), 293–298.

Elmer, M. I. (1987). *Intercultural effectiveness: Development of an intercultural competency scale* (Unpublished doctoral dissertation). Michigan State University, East Lansing, MI.

Engle, R. L., & Crowne, K. A. (2014).The impact of international experience on cultural intelligence: An application of contact theory in a structured short-term programme. *Human Resource Development International, 17*(1), 30–46.

Engle, R. L., & Crowne, K. A. (2014).The impact of international experience on cultural intelligence: An application of contact theory in a structured short-term programme. *Human Resource Development International, 17*(1), 30–46.

Erez, M., & Gati, E. (2004). A dynamic, multi-level model of culture: From the micro level of the individual to the macro level of a global culture. *Applied Psychology, 53*, 583–598.

Erez, M., Lisak, A., Harush, R., Glikson, E., Nouri, R., & Shokef, E. (2013). Going global: Developing management students' cultural intelligence and global identity in virtual culturally diverse teams. *Academy of Management Learning & Education, 12*, 330–355

Fantini, A. E., & Tirmizi, A. (2006). *Exploring and assessing intercultural competence*.World Learning Publications, Paper 1. Brattleboro VT: SIT Digital Collections. Retrieved from http://digitalcollections.sit.edu/worldlearning_publications/1

Fuertes, J. N. (2000). Factor structure and short formof theMiville-Guzman Universality-Diversity Scale.*Measurement and Evaluation in Counseling and Development, 33*(3), 157–169.

Global Leadership Excellence. (2010). RetrievedMay 8, 2015, from http:www.globallycompetent.com sectors/GCAAeducation.html. Goldberg, L. R. (1999). A broad-bandwidth, public domain personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe, 7*, 7–28.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality Psychology in Europe, 7*, 7–28.

Graf,A.,&Mertesacker,M. (2009). Intercultural training: Sixmeasures assessing training needs. *Journal of European Industrial Training, 33*(6), 539–558.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41. doi:10.1037/a0015575

Hammer, M. R. (2005). *Assessment of the impact of the AFS study abroad experience. Executive summary: Overall findings*. Retrieved from http://idiinventory.com/wp-content/uploads/2014/02/afs_study.pdf

Hammer, M. R. (2011). Additional cross-cultural validity testing of the Intercultural Development Inventory. *International Journal of Intercultural Relations, 35*(4), 474–487.

Hammer, M. R. (2012). The Intercultural Development Inventory: A new frontier in assessment and development of intercultural competence. In M. Vande Berg, R. M. Paige, & K. H. Lou (Eds.), *Student learning abroad* (pp. 115–136). Sterling, VA: Stylus.

Hammer, M. R., Bennett, M. J., &Wiseman, R. (2003). Measuring intercultural sensitivity: The Intercultural Development Inventory. *International Journal of Intercultural Relations, 27*(4), 421–443.

Hammer, M. R., Wiseman, R. L., Rasmussen, J. L., & Bruschke, J. C. (1998)

Hanvey, R. G. (1982). An attainable global perspective. Theory Into Practice, 21(3), 162–167. doi:10.1080/00405848209543001

Harrison, J. K. (1992). Individual and combined effects of behavior modeling and the culture assimilator in cross-cultural management training. *Journal of Applied Psychology, 77*, 952–962.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. NewYork, NY: SAGE.

Hunter, W. D., White, G. P., & Godbey, G. C. (2006). What does it mean to be globally competent? *Journal of Studies in International Education, 10*(3), 267–285.

Hofstede, G. (2010).The GLOBE debate: Back to relevance. Journal of International Business Studies, 41(8), 1339-1346.

Ingulsrud, J. E., Kai, K., Kadowaki, S., Kurobane, S., & Shiobara, M. (2002). The assessment of cross-cultural experience: Measurin awareness through critical text analysis. *International Journal of Intercultural Relations, 26*(5), 473–491.

Isley, E. B., Shealy, C. N., Crandall,K. S., Sivo, S.A.,&Reifsteck, J. B. (1999,August). *Relevance of the "BEVI" for research in developmental psychopathology.* Poster session presented at the meeting of the American Psychological Association, Boston, MA.

Jacobson,W., Sleicher,D., &Maureen, B. (1999). Portfolio assessment of intercultural competence. *International Journal of Intercultural Relations, 23*(3), 467–492.

Kelley, C., & Meyers, J. (1995). Cross-Cultural Adaptability Inventory. Minneapolis, MN: National Computer Systems.

Klein, J. (1995). Intelligence and cross-cultural sensitivity. Psychology: A Journal of Human Behavior, 32(1), 31–32.

Koban, L., & Pourtois, G. (2014). Brain systems underlying the affective and social monitoring of actions: An integrative review. *Neuroscience & Biobehavioral Reviews*, *46*, 71–84.

Koester, J., & Olebe, M. (1989).The behavioral assessment scale for intercultural communication effectiveness. *International Journal of Intercultural Relations, 12*(3), 233–246. doi:10.1016/0147-1767(88)90017-X

Kupka, B. (2008). *Creation of an instrument to assess intercultural communication competence for strategic international human resource management* (Unpublished doctoral dissertation). University of Otago, Otago, New Zealand.

Kupka, B., & Everett, A. M. (2008). Moscow on the Hudson: Assessing expatriates' affective fit in host cultures with the intercultural communication affinity scale. *European Journal of International Management*, *2*(3), 234–249.

Leung, K., Ang, S.,&Tan,M. L. (2014). Intercultural competence. *Annual Review of Organizational Psychology and Organization Behavior, 1*(1), 489–519.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426–441.

Lin, Y. C., Chen, A. S. Y., & Song, Y. C. (2012). Does your intelligence help to survive in a foreign jungle? The effects of cultural intelligence and emotional intelligence on cross-cultural adjustment . *International Journal of Intercultural Relations, 36*, 541–552.

Lisak, A., & Erez, M. (2015). Leadership emergence in multicultural teams:The power of global characteristics. *Journal ofWorld Business*, *50*(1), 3–14.

Matsumoto, D., LeRoux, J. A., Ratzlaff, C., Tatani, H., Uchida, H., Kim, C., & Araki, S. (2001). Development and validation of a measure of intercultural adjustment potential in Japanese sojourners: The Intercultural Adjustment Potential Scale (ICAPS). *International Journal of Intercultural Relations, 25*, 483–510.

Messick, S. (1995).Validity of psychological assessment:Validation of inferences frompersons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749.

Messner, W., & Schؘfer, N. (2012). *The ICCA facilitator's manual: Intercultural Communication and Collaboration Appraisal*. London, England: GloBus Research

Miville, M. L., Gelso, C. J., Pannu, R., Liu, W., Touradji, P., Holloway, P., & Fuertes, J. (1999). Appreciating similarities and valuing differences: The Miville-Guzman Universality Diversity Scale. *Journal of Counseling Psychology, 46*, 291–307

.

Morrell,D. L., Ravlin, E. C., Ramsey, J. R.,&Ward,A.K. (2013). Past experience, cultural intelligence, and satisfaction with international business studies. *Journal of Teaching in International Business, 24*, 31– 43.

Nguyen, N. T., Biderman,M. D., &McNary, L. D. (2010). A validation study of the Cross-Cultural Adaptability Inventory. *International Journal of Training and Development, 14*(2), 112–129.

Olebe, M., & Koester, J. (1989). Exploring the cross-cultural equivalence of the Behavioral Assessment Scale for Intercultural Communication. *International Journal of Intercultural Relations, 13*(3), 333–347. doi:10.1016/0147-1767(89)90016-3

Pruegger, V. J., & Rogers, T. B. (1993). Development of a scale to measure cross-cultural sensitivity in the Canadian context. Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement, 25(4), 615–6

Remhof, S., Gunkel,M., & Schlaegel, C. (2013).Working in the "Global Village":The influence of cultural intelligence on the intention to work abroad. *Zeitschrift für Personalforschung/German Journal of Human Resource Management, 27*(3), 224–250.

Rios, J. A., &Wells, C. S. (2014). Validity evidence based on internal structure. Psicothema, 26(1), 108–116.

Ruben, B. D. (1976). Assessing communication competency for intercultural adaptation. Group and Organization Studies, 1, 334–354.

Sampson, D. L., & Smith, H. P. (1957. A scale to measure world-minded attitudes.The Journal of Social Psychology, 45(1), 99–106.

Schmitz, J., Tarter, L., & Sine, J. (2012). *Understanding the cultural orientations approach: An overview of the development and updates to the COA*. Retrieved from

https://www.culturalorientations.com/SiteData/docs/ArticleUnd/616d3a22b5d5d472/Article%20-

Shealy, C. N., Burdell, L. L., Sivo, S. A., Davino, D. F., & Hayes, D. J. (1999, August).*Men, masculinity, and Scale 10 (Gender Stereotypes) of the "BEVI."* Poster session presented at the meeting of the American Psychological Association, Boston, MA.

Shealy, C. N., Sears, J. L., Sivo, S. A., Alessandria, K. P., & Isley, E. B. (1999, August). *Intercultural psychology and Scale 3 (Sociocultural Closure) of the "BEVI."* Poster session presented at the meeting of the American Psychological Association, Boston, MA.

Shirts, G. R. (1977). *BaFa' BaFa' – A cross-cultural simulation*. La Jolla, CA: Simile II.

Shokef, E., & Erez, M. (2006). Global work culture and global identity, as a platform for a shared understanding in multicultural teams. *National Culture and Groups, 9*, 325–352.

Shokef, E.,&Erez, M. (2008). Cultural intelligence and global identity inmulticultural teams. In S.
Ang&L.VanDyne (Eds.), *Handbook of cultural intelligence:Theory, measurement, and applications* (pp. 177–191). Abington, England: M. E. Sharpe.

Skinner, M. (2002).The renaissance of unconventional warfare as an SF mission. *SpecialWarfare, 15*, 16–22.

Smith, D. E., & Mitry, D. J. (2008). Benefits of study abroad and creating opportunities:The case for short-term programs. *Journal of Research in Innovative Teaching, 1*(1), 236–246.

Stephan,W. G., & Stephan, C. (1985). Intergroup anxiety. *Journal of Social Issues, 41*, 157–176.

Stevens, M., Bird, A., Mendenhall, M. E., & Oddou, G. (2014). Measuring global leader intercultural competency: Development and validation of the Global Competencies Inventory (GCI). *Advances in Global Leadership, 8*, 115–154.

Strubler, D., Agarwal, A., Park, S., & Elmer,M. (2011). From cognition to behavior: A cross cultural study for global business effectiveness. *Journal of International Business Research, 10*, 35–46.

Thomas,D. C., & Lazarova,M. B. (2006). Expatriate adjustment and performance: A critical review. In G. K. Stahl & I. Bjorkman (Eds.), *Handbook of research in international human resource management* (pp. 247–265). Northampton, MA: Edward Elgar.

Thomas, D. C., Stahl, G., Ravlin, S., Poelmans, A., Pekerti, M., Maznevski, M., … Brislin, R. (2015). Development of the Cultural Intelligence Assessment. *Advances in Global Leadership, 8*, 155–178.

Van der Zee, K. I., Atsma, N., & Brodbeck, F. (2004).The influence of social identity and personality on outcomes of cultural diversity in teams. *Journal of Cross-Cultural Psychology, 35*(3), 283–303.

Van der Zee, K. I., & Van Oudenhoven, J. P. (2000).The Multicultural Personality Questionnaire: A multidimensional instrument of multicultural effectiveness. *European Journal of Personality, 14*(4),
291–309

Varela, O., & Gatlin-Watts, R. (2013).The development of the globalmanager: An empirical study on the role of academic international sojourns. *Academy of Management Learning & Education, 13*, 187–207

Wang, Y. W., Davidson, M. M., Yakushko, O. F., Savoy, H. B., Tan, J. A., & Bleier, J. K. (2003). The scale of ethnocultural empathy: Development, validation, and reliability. *Journal of Counseling Psychology, 50*(2), 221–234.

Ward, C., & Kennedy, A. (1999).The measurement of sociocultural adaptation. *International Journal of Intercultural Relations, 23*(4), 659–677.

Ward, C.,Wilson, J., & Fischer, R. (2011). Assessing the predictive validity of cultural intelligence over time. *Personality and Individual Differences, 51*, 138–142.

Whetzel, D. L., &McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*, 188–202.

Wiseman, R. L., Hammer, M. R., & Nishida, H. (1989). Predictors of intercultural communication competence. *International Journal of Intercultural Relations, 13*(3), 349–370.