Assessing Intercultural Competence in Higher Education: Existing Research and Future Directions

Richard L. Griffith

Leah Wolfeld

Brigitte K. Armon

Joseph Rios

Ou Lydia Liu

December 2016

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

Assessing Intercultural Competence in Higher Education: Existing Research and Future Directions

Richard L. Griffith,[1] Leah Wolfeld,[1] Brigitte K. Armon,[1] Joseph Rios,[2] & Ou Lydia Liu[2]

1 Institute for Cross Cultural Management, Florida Institute of Technology, Melbourne, FL
2 Educational Testing Service, Princeton, NJ

The modern wave of globalization has created a demand for increased intercultural competence (ICC) in college graduates who will soon enter the 21st-century workforce. Despite the wide attention to the concepts and assessment of ICC, few assessments meet the standards for a next-generation assessment in areas of construct clarity, innovative item types, response processes, and validity evidence. The objectives of this report are to identify current conceptualizations of ICC, review existing assessments and their validity evidence, propose a new framework for a next-generation ICC assessment, and discuss key assessment considerations. To summarize, we found the current state of the literature to be murky in terms of the clarity of the ICC construct. Definitions of the construct vary considerably as to whether it is a trait, skill, or performance outcome. In addition, current measurements of ICC overly rely on self-report methods, which have a number of flaws that result in less than optimal assessment. In this paper, we propose a new framework based on a model of the social thinking process developed by Grossman and colleagues that describes the knowledge, skills, and abilities that promote success in complex social situations. From this social process model, as well as Earley and Peterson's definition of ICC (a person's capability to gather, interpret, and act upon these radically different cues to function effectively across cultural settings or in a multicultural situation), three stages are developed: approach, analyze, and act. Guided by this framework, we discuss assessment considerations such as innovative task types and multiple response formats to help translate the framework to an assessment of ICC.

Keywords Intercultural competence; measurement; cross-cultural competence; global competence; international higher education

The modern wave of globalization, having long overtaken the business sector, economics, technology, and transportation, has come to higher education. To compete in the global arena — and, therefore, solicit international student revenue, attract high-potential students, and produce effective university ambassadors for increased brand recognition — university administrators must demonstrate that their institution prepares graduates appropriately for the global workforce. In the last 8 years, the United States witnessed a 56% increase of international students studying in higher education institutions, resulting in 886,052 additional students for the 2013–2014 school year, which generated 30.5 billion dollars for the U.S. economy (Institute of International Education, 2015) and created 373,000 jobs (NAFSA: Association of International Educators, 2016). For years, prestigious programs such as the Fulbright Program have been sending students and scholars around the world to higher education institutions to facilitate mutual understanding across countries (Bureau of Educational and Cultural Affairs, 2013). Further, 273,996 U.S. students enrolled in higher education studied abroad in the 2012–2013 academic year (Institute of International Education, 2015). Thus, increased internationalization in higher education institutions alone demands that university students develop intercultural competence (ICC) in order to interact successfully with diverse peers and professors and maximize their collegiate experience.

Being able to communicate and work effectively across cultures has also been identified as a desirable capability by various organizations with global missions (Bikson, Treverton, Moini, & Lindstrom, 2003) and even more important to potential employers than an undergraduate major; in fact, 78% of surveyed employers stressed the importance of all students gaining intercultural skills (Hart Research Associates, 2015). Unsurprisingly, ICC has been identified as an essential student learning outcome in higher education (Association of American Colleges and Universities, 2011). Accordingly, higher education institutions in the United States and abroad are increasingly concerned with preparing students to be competitive contributors in the global economy as well as remaining competitive in regard to international education and other internationalization efforts (e.g., exchange programs, study abroad experiences, and marketing targeted toward

Corresponding author: Ou Lydia Liu, E-mail: LLiu@ets.org

international students; De Haan, 2014; Scott, 2000). If higher education institutions are to remain relevant, they must take charge of their internationalization and produce graduates who will excel in the global work arena (e.g., Fellows, Goedde, & Schwichtenberg, 2014). Meeting the challenge of producing culturally competent graduates requires the tracking of student development of ICC; however, the existing challenges of measuring ICC complicate tracking initiatives.

Although some higher education institutions recognize the importance of measuring their students' ICC, this recognition has only recently expanded beyond assessing study abroad programs. For instance, the Fund for the Improvement of Postsecondary Education (FIPSE) program, through the U.S. Department of Education, has developed an international learning outcomes ranking document to help institutions prioritize and assess components of ICC. (Its website may be found at http://www2.ed.gov/about/offices/list/ope/fipse/index.html). Another initiative, At Home in the World: Educating for Global Connections and Local Commitments (AHITW), sponsored by the American Council on Education (ACE), highlights the need to include assessment as part of developing student and institutional ICC (ACE, 2016). Thus, the awareness of the benefits of higher education institutions assessing ICC among all students, not just those who participate in study abroad or exchange programs, is spreading. However, as will be discussed in detail in this report, many of the measures available to university administrators are self-report measures, some with inadequate evidence of reliability and validity.

Given that higher education institutions have identified ICC to be a valuable student outcome and a marketable indicator of student and overall institutional success, it is imperative to develop valid and reliable measures of ICC in the context of higher education. Such an initiative would facilitate assessment of two areas: the capability of institutions to graduate interculturally competent students and the quality of various educational experiences in terms of student development. The purpose of this report is to explore the possibility and utility of assessing ICC for students in higher education. To this end, we review current definitions, existing assessments, and challenges for measuring this multidimensional construct. We then propose a theoretical model of ICC to guide the design of an assessment that captures the complexity of the construct while avoiding its common measurement pitfalls. After describing the model, we then describe several measurement considerations, including task type, response format, and the need for more advanced assessment techniques.

Current State of Assessments, Research, and Challenges

Definitions of Intercultural Competence in Higher Education

A review of the literature (see Appendix for a description of the literature search process) revealed a multitude of definitions of ICC. The ICC definitions (Table 1) used in the higher education literature tend to be associated with models used in education, training, and research. These models fall into five categories: compositional, co-orientational, developmental, adaptational, and causal (Spitzberg & Changnon, 2009). Compositional models (e.g., Deardorff, 2006; W. D. Hunter, White, & Godbey, 2006; Ting-Toomey & Kurogi, 1998) merely describe the characteristics (knowledge, skills, and attitudes) of ICC. Co-orientational models (e.g., Fantini, 1995; Kupka, 2008; Rathje, 2007) tend to describe the components or process of a successful intercultural interaction. Developmental models describe ICC in terms of individual development over time (e.g., Bennett, 1986; P. M. King & Baxter Magolda, 2005). Adaptational models (e.g., J. W. Berry, Kim, Power, Young, & Bujaki, 1989; Gallois, Franklyn-Stokes, Giles, & Coupland, 1988) combine the developmental components of the aforementioned models and present them in an interactional context of adapting to a foreign culture. Finally, causal path models (e.g., Arasaratnam, 2008; Deardorff, 2006; D. A. Griffith & Harvey, 2000; Hammer, Wiseman, Rasmussen, & Bruschke, 1998) attempt to integrate the characteristics of compositional models and situate them in an interaction in which variables influence each other to predict ICC.

A recent review of ICC focusing on research across multiple contexts (Leung, Ang, & Tan, 2014) presented another system of grouping ICC models. This system differentiates between models that include intercultural traits, intercultural attitudes and worldviews, and intercultural capabilities, or some mix thereof. The term *intercultural traits* refers to stable personality traits that drive likely behavior, and they commonly include openness to experience and tolerance for ambiguity. The term *intercultural attitudes and worldviews* refers to constructs involving the perception and evaluation of information from outside an individual's own culture. Lastly, the term *intercultural capabilities* refers to anything that a person can do, think, or know that will allow him or her to interact successfully in an intercultural situation.

Neither scholars in the field of ICC nor higher education administrators have reached a consensus regarding the definition of ICC and its underlying dimensions. For example, in a recent study, administrators from 24 U.S. postsecondary

Table 1 Descriptions of Intercultural Competence (ICC) and Relevant Constructs from Conceptual Models in Higher Education and Business Research (in Chronological Order)

Source(s)	Construct(s)/dimensions	Description	Model type
Bennett (1986)	Intercultural sensitivity	"the way people construe cultural difference and … the varying kinds of experience that accompany these constructions" (Bennett, 1993, p. 24) Development of intercultural sensitivity through six stages: denial, defense/reversal, minimization, acceptance, adaptation, and integration.	Developmental
Gallois et al. (1988)	Intercultural communicative accommodation: situational factors, individual factors, and encoding/decoding factors	Interacting individuals adjust their communication styles to match the other individual's style. Competence is judged both within and between groups.	Adaptational
J. W. Berry et al. (1989)	Acculturation: integration, assimilation, separation/segregation, marginalization	Views toward adapting to a foreign culture and retaining one's cultural identity can be both orthogonal and dichotomous.	Adaptational
Imahori and Lanigan (1989)[b]	Intercultural competence: intercultural/communication effectiveness, relational validation, satisfaction, intimacy, commitment, stability, and uncertainty reduction.	Relationship subfactors index, competent intercultural interaction between a sojourner and a host–national	Casual
Y. Y. Kim (2000)[a]	Host communication competence	One's adaptive capacity to suspend/modify old cultural ways, learn/accommodate to new cultural ways, and creatively manage dynamics of cultural difference/unfamiliarity and accompanying stress	Adaptational
Paige (1993) and Paige and Goode (2009)[a]	Intercultural learning: self as a cultural being, elements of culture, culture-specific learning, culture-general learning, learning about learning	Ability to effectively function in an intercultural situation abroad over time	Co-orientational
Gudykunst et al. (1994) and Pusch (1994)[a]	Global competence: mindfulness, cognitive flexibility, tolerance for ambiguity, behavioral flexibility, cross-cultural empathy	Motivation, knowledge, and skills make up global competence	Co-orientational
Lambert (1994)[a,b]	Global competence	World knowledge, foreign language proficiency, cultural empathy, approval of foreign people and cultures, ability to practice one's profession in an international setting	Compositional
Fantini (1995)[b] and Fantini, Arias-Galicia, and Guay (2001)	Intercultural communicative competence	"1) The ability to develop and maintain relationships, 2) the ability to communicate effectively and appropriately with minimal loss or distortion, and 3) the ability to attain compliance and obtain cooperation with others" (Fantini et al., 2001, p. 27)	Co-orientational
Chen and Starosta (1996)[a,b]	Intercultural communication competence: intercultural sensitivity (affective process), intercultural awareness (cognitive process), and verbal/nonverbal skills (Fantini et al., 2001). May include motivation dimension (Yamaguchi and Wiseman, 2001).	Ability to effectively and appropriately execute communication behaviors in a culturally diverse environment	Compositional

Table 1 Continued

Source(s)	Construct(s)/dimensions	Description	Model type
Byram (1997)[a,b]	Communicative competence (CC)	"Knowledge of others; knowledge of self; skills to interpret and relate; skills to discover and/or to interact; valuing others' values, beliefs, and behaviors; and relativizing one's self. Linguistic competence plays a key role." (Byram, 1997, p. 34)	Co-orientational
Fennes and Hapgood (1997)[a]	Intercultural learning	The expandability, flexibility, and adaptability of one's frame of reference/filter	Compositional
Howard-Hamilton, Richardson, and Shuford (1998)	ICC: knowledge, attitudes, and skills	Competence components consist of knowledge, attitudes, and skills across three levels: awareness, understanding, and appreciation of another culture.	Compositional
Hammer et al. (1998)[b]	ICC: interpersonal/intergroup saliencies, communication message exchange, host contact conditions, attributional confidence, anxiety reduction, satisfaction	Satisfying intercultural interactions are mediated by conditions, strategies, and saliencies that lead to greater attribution confidence and reduction of uncertainty (anxiety).	Casual
Ting-Toomey and Kurogi (1998)	ICC: knowledge, mindfulness, interaction skills, and facework competence criteria	Cognitive and behavioral abilities are predicted to increase the likelihood of positive (appropriate, effective, mutually satisfying, and mutually adaptive) intercultural interactions	Compositional
Ting-Toomey (1999)	ICC: personal, interpersonal, and system-level outcomes	The ability to manage changes in the self and the environment brought about by individual, interpersonal, and systemic influences during intercultural encounters	Casual
D. A. Griffith and Harvey (2000)	Intercultural communication competence: cultural understanding, cultural interaction, communication interaction, relationship quality	A component in a network of intercultural constructs that collectively can be judged by the criterion of relationship quality; cultural understanding and intercultural communication competence directly predict relationship quality.	Casual
Koester and Olebe (1989)*	Intercultural communication effectiveness: display of respect, orientation to knowledge, empathy, interaction management, task role behavior, relational role behavior, tolerance for ambiguity, and interaction posture	Behaviors that a nonexpert, nonnative English speaker can reliably assess as effective or not in a cross-cultural setting	Compositional
Lustig and Koester (2003)[a]	ICC	Not comprised of individual traits or characteristics but rather the characteristic of the association between individuals. Dependent on the relationships and situations within which the interaction occurs. No prescriptive set of characteristics guarantees competence in all intercultural situations: "social judgment that people make about others." (Lustig & Koester, 2003, pp. 64–65).	Co-orientational
Deardorff (2004, 2006)[L]	ICC: requisite attitudes, knowledge and comprehension, skills, desired internal outcomes, desired external outcomes	"the ability to communicate effectively and appropriately in intercultural situations based on one's intercultural knowledge, skills, and attitudes" (Deardorff, 2004, p. 194).	Compositional/Causal Path

Table 1 Continued

Source(s)	Construct(s)/dimensions	Description	Model type
P. M. King and Baxter Magolda (2005)	Intercultural maturity: cognitive, intrapersonal, interpersonal components across three levels (initial, intermediate, and mature development)	Through ongoing study, observation, and interaction with individuals from another culture, one can develop greater intercultural awareness and sensitivity.	Developmental
Navas et al. (2005)	Relative adaptation: real and ideal attitudes held and strategies implemented by immigrant and native groups across situations	Extent of competence depends on the alignment between the strategies actually used by one group and the preferences of the other group.	Adaptational
W. D. Hunter et al. (2006)	Global competencies model: (inner box) recognition of others/differences, openness to new experiences and diversity, nonjudgmental stance; (middle box) understanding of world history and globalization; (outer box) identification of cultural difference to compete globally, effective participation (socially and in business) across cultures, collaboration across cultures, and ability to assess intercultural performance	"a person should attempt to understand his or her own cultural box before stepping unto someone else's" (p. 279).	Compositional
Rathje (2007)	ICC	"transformation of intercultural interaction into culture itself" (p. 263); the coproduction of culture, not just the reflection of common cultural identities	Co-orientational
Arasaratnam (2008)	Intercultural communication competence: cultural empathy, experience, interaction involvement (conversation awareness), global attitude, and motivation	Intercultural communication competence is a direct function of cultural empathy. Motivation for competent communication is influenced by experience, interaction involvement, and one's global attitude, as well as prior experience with intercultural communication.	Casual
Kupka (2008)	ICC: basic human needs, culture A/B conceptas and perceptas, noise	"impression management that allows members of different cultural systems to be aware of their cultural identity and cultural differences, and to interact effectively and appropriately with each other in diverse contexts by agreeing on the meaning of diverse symbol systems with the result of mutually satisfying relationships" (p. 16)	Co-orientational

Note. In the first column, the source(s) from which the definition of ICC was retrieved is listed. The name of the relevant construct(s) and any dimensions of the construct are listed in the second column followed by a description of the definition given for the construct(s). The last column specifies the type of model in which each definition was used, per Spitzberg and Changnon's (2009) typology of compositional, co-orientational, developmental, adaptational, and causal models.
[a]Denotes inclusion in Deardorff (2004), in which HEI administrators rated the definitions. [b]Denotes models that are language-focused.

institutions rated nine definitions of ICC on a 4-point scale (4 = *highly applicable* and 1 = *not applicable*; Deardorff, 2006). The results demonstrated that Byram's (1997) definition of ICC, which focuses heavily on language proficiency, was the highest rated (*M* = 3.5), followed by Lambert's (1994) definition (*M* = 3.3), which highlights task accomplishment in the global context (see Table 1; Deardorff, 2006). Responses from administrators also revealed that similar yet distinctive terms were being used to discuss this construct, including c*ross-cultural competence, global competence, intercultural competence*, and *global citizenship* (Deardorff, 2006, p. 247), and confirmed the need for a general definition that could be used across student populations and contexts.

In an effort to find a widely agreed-upon definition, the same researchers identified three prevalent themes across definitions generated by individual institutions, including "the awareness, valuing, and understanding of cultural differences; experiencing other cultures; and self-awareness of one's own culture" (Deardorff, 2006, p. 247). In the same study, a group of 23 international scholars rated the same nine definitions; on average, Deardorff's (2004) definition of ICC as "the ability to communicate effectively and appropriately in intercultural situations based on one's intercultural knowledge, skills, and attitudes" (p. 194) was the highest rated. In addition, the scholars generated definitions and specific elements of ICC. Seven definitions and 22 elements were agreed upon by 80% (16 out of 23) of the group, with only one element, understanding of others' world views, receiving 100% agreement from the raters. Although this particular study may have achieved some clarity and alignment on defining ICC in the higher education context, further agreement remains elusive, in part due to the existence of multiple alternative models (e.g., Fantini & Tirmizi, 2006). In addition, abstract, complex phenomena are often better defined through the process of measurement; however, many of the existing theories and models of ICC are not clarified through validated measurement. Therefore, the framework presented in this paper incorporates both theoretical and measurement considerations.

Discrepancies in Dimensional Models of Intercultural Competence

This variability in content of ICC models and dimensions presents several challenges. First, it reduces the conceptual clarity of the construct itself, as some models include as core components factors that are excluded or treated as antecedents in other models. For example, *tolerance for ambiguity*, which refers to the ability to make progress despite high levels of uncertainty (Bird, Mendenhall, Stevens, & Oddou, 2010), is included in some definitions and measures (e.g., Deardorff, 2006; Gudykunst, 2003) but excluded in others (e.g., Byram, 1997). Second, in addition to reducing the conceptual clarity of ICC, these discrepancies complicate the specification of ICC's nomological network (i.e., the constructs theorized or empirically related to ICC). Specifically, existing literature has yet to distinguish constructs belonging in the ICC framework from its correlates. Constructs such as global mindedness, broadmindedness, cosmopolitanism, and global identity provide prime examples. Because the definitions of these constructs are imprecise and vary considerably, it can be challenging to determine which of these constructs reflect a subfacet of ICC and which constitute a part of its nomological network. Third, several constructs demonstrate significant overlap with ICC—including the global leadership construct that has recently received much attention (Bird et al., 2010). The existing literature has yet to fully delineate where one ends and another begins (Bücker & Poutsma, 2010). In sum, establishing construct validity for ICC is a less straightforward task than it is for other, less complex concepts. Any new model of ICC attempting to address these concerns should meet the following criteria: (a) provide specific definitions of the overall construct and its subdimensions, (b) include both cognitive and noncognitive components, and (c) clarify the relationship between subdimensions. To date, many of the models of ICC do not meet the above criteria. Although many models are multidimensional in nature, models focusing only on attitudes (or attitudes and cognitions) are prevalent, thereby lacking the focus on the behavioral or performance-relevant component of ICC. Other scales rely on weak definitions or do not clarify the relationship among subdimensions.

Malleability of Intercultural Competence in the Higher Education Context

Some evidence suggests that ICC is a malleable skill and that higher education experiences influence the development of these competencies for both educators and students (e.g., Eisenberg et al., 2013). Most intercultural education research focuses on best practices to train K–12 teachers to work effectively with diverse student populations (DeJaeghere & Cao, 2009; DeJaeghere & Zhang, 2008; Teräs & Lasonen, 2013). Similarly, the research on ICC in higher education focuses on training international education professionals, which include roles such as collegiate language instructors, study abroad and international student advisors, faculty members, and other professionals supporting international educational exchange programs (Paige & Goode, 2009, p. 333).

A small body of research focuses on student development (e.g., Conway, 2008; DeJaeghere & Zhang, 2008; Fischer, 2011; Hao, 2012; Jauregui, 2013; Kahr-Gottlieb & Papst, 2013; Kaufmann, Englezou, & García-Gallego, 2014; Zhang, 2012). These studies indicate that ICC may be improved with training, including study abroad programs (e.g., Engle & Crowne, 2014) and intercultural business courses (e.g., Eisenberg et al., 2013; Rosenblatt, Worthley, & MacNab, 2013). Despite the prevalence of the training and activities surrounding this area, the empirical evidence documenting their effectiveness is nascent, precluding strong conclusions on the best ways to improve ICC. However, initial evidence suggests that ICC is a malleable construct and that higher education may improve students' ICC (e.g., Williams, 2005).

Existing Assessments of Intercultural Competence

Multidimensional Nature of Intercultural Competence Assessments

Corresponding to the wide-ranging models and conceptualizations of ICC reviewed in the previous section, existing assessments of ICC vary in the number of constituent constructs and dimensions to be measured. Some scholars operationalize ICC as unidimensional and measure it with all items loading onto one factor (e.g., Global Perspective Survey; Hanvey, 1982), although others argue that ICC is multidimensional, including dimensions such as approachableness, intercultural receptivity, positive orientation, forthrightness, social openness, enterprise, respectfulness, flexibility, perseverance, cultural perspectivism, venturesome, and social confidence (e.g., Intercultural Competency Scale; Elmer, 1987). Table 2 presents existing assessments used to measure ICC in higher education and business contexts, including those reviewed by Fantini (2009) but excluding those that measure language ability.

The ICC instruments reviewed in this study vary substantially in terms of how they define the ICC dimensions. Some assessments conceptualize ICC as having separate, broad dimensions such as cognitive, interpersonal, intrapersonal, metacognitive, affective, motivational, and behavioral, but others use terms such as knowledge, skills, attitudes, processes, and awareness. Despite their differences in categorization, ICC instruments have overlapping dimensions. For example, the dimensions of openness, flexibility, and empathy appear in multiple assessments. Additionally, several models nest specific competencies and traits within subdimensions (e.g., the cultural intelligence construct divides its competencies into metacognitive, cognitive, behavioral, and motivational domains; Earley & Ang, 2003).

Assessment Formats

Currently, two predominant assessment formats are used to measure ICC: surveys and portfolio assessments. All of the instruments reviewed in Table 2 are administered as surveys ranging in length from nine items (i.e., Global Perspective Survey; Hanvey, 1982) to over 160 items (i.e., Intercultural Communication and Collaboration Appraisal; Messner & Schäfer, 2012). Typically, these surveys are delivered through an online format, though some assessments (e.g., Intercultural Development Inventory; Hammer, Bennett, & Wiseman, 2003) are also offered in a paper and pencil format. This article reviewed only ICC assessments that exclusively used selected-response items.

In addition to surveys, portfolios that include constructed-response items are also used to assess ICC in higher education. A portfolio assessment is a collection of materials produced either by an individual over time or scores from various assessments or both. Currently, no standard portfolio assessment exists, meaning that the content, platform (paper vs. digital), and scoring method vary across institutions, studies (e.g., Ingulsrud, Kai, Kadowaki, Kurobane, & Shiobara, 2002; Jacobson, Sleicher, & Maureen, 1999), and contexts (e.g., foreign language courses, study abroad experiences, general education). This deficit can be viewed as an advantage. Portfolios are able to capture context-specific skills (e.g., writing business letters for a local business owner in a third-world country) and the development of those skills over time. Thus, ICC is captured through the collection of work products from different time points in a student's career (e.g., before, during, and after an experience abroad; Ingulsrud et al., 2002; Jacobson et al., 1999).

Some higher education institutions worldwide use digital portfolios. For example, Alliant International University uses a digital portfolio format to assess ICC in its study abroad students. Clemson University also uses a digital portfolio and requires all students to provide evidence of cross-cultural awareness as a universal general education requirement, regardless of participation in programs abroad. Evidence of cross-cultural awareness, which Clemson University (2016) defines as "the ability to critically compare and contrast world cultures in historical and/or contemporary contexts" (bullet 2), is demonstrated in digital portfolios through the inclusion of writing samples. Although digital portfolios have the

Table 2 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/adaptation, 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.

Table 2 Continued

Test	Developed (year)	Format	Delivery	Forms and items	Themes/topics
Multicultural Personality Questionnaire (MPQ)	Van der Zee and Van Oudenhoven (2000)	Self-report; 5-point Likert scale (*not at all applicable* to *totally applicable*)	Online survey	78 items	Measures multicultural effectiveness through five subscales: cultural empathy, 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 growth.
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 tool.
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, and culture.
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 Competency Scale (ICS)	Elmer (1987)	Self-report; response scale unknown	Online survey	45 items	Measures intercultural effectiveness through 12 factors, such as approachable, intercultural receptivity, positive orientation, forthrightness, social openness, enterprise, shows respect, flexibility, perseverance, cultural perspectivism, venturesome, and social confidence.

Table 2 Continued

Test	Developed (year)	Format	Delivery	Forms and items	Themes/topics
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 groups
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 of contact, relativism appreciation, and comfort with difference.
Cross-Cultural World-Mindedness Scale (CCWM)	Der-Karabetian (1992)	Self-report; response scale unknown	?	?	Measures worldmindedness.
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	Paper and pencil	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 Sensitivity Scale (CCSS)	Pruegger and Rogers (1993)	Self-report; 6-point Likert scale; *strongly disagree* to *strongly agree*	Paper and pencil	24 items total (two equivalent forms with 12 items each)	Measures the valuation and tolerance of different cultures.

Table 2 Continued

Test	Developed (year)	Format	Delivery	Forms and items	Themes/topics
Intercultural Communication Competence (ICC)	Arasaratnam and Doerfel (2005) and Arasaratnam (2009)	Self-report: 7-point Likert scale; *strongly disagree to strongly agree*	Paper and pencil	10 items; 3–4 items for each dimension	Cognitive, affective, and behavioral dimensions of intercultural communication competence
Intercultural Sensitivity Inventory (ICSI)	Bhawuk and Brislin (1992)	Self-report: 7-point Likert scale; *very strongly disagree to very strongly agree*	Paper and pencil	46 items; 14–16 items per subscale; individualism versus collectivism are asked in relation to own or other culture	Measures individualism versus collectivism and flexibility/open-mindedness.
Global Competencies Inventory	Kozai Group; Bird et al. (2002) Stevens, Bird, Mendenhall, and Oddou (2014)	Self-report: 5-point Likert scale; *strongly disagree to strongly 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 Intelligence	Ascalon et al. (2008)	SJT: 4 response options	?	14 scenarios; replies vary across ethnocentric–nonethnocentric and empathetic–nonempathetic	Measures knowledge, skills, and other characteristics that promote successful social interaction in cross-cultural interactions.
Cultural Intelligence Assessment	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.
Nonverbal Communication Competence Scale (NVCCS)	Kupka and Everett (2008)	Self-report; anchors unknown	Paper and pencil	5 items	Measures the degree of knowledge that is essential to recognize nonverbal 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 evaluated.

capability to include other work products such as audio and video recordings of intercultural communication (Deardorff, 2009), institutions that actually request such products have not been identified.

As with all assessments, their format largely depends on the intended purpose of the assessment. Although ICC experts suggest that more than one methodology (i.e., both qualitative and quantitative methods) should be used to measure ICC (Deardorff, 2006; Fantini, 2009), assessing ICC for higher education institutions to provide benchmark information about students' ICC requires a format that allows meaningful comparisons of individuals and groups of examinees. For this purpose, portfolios may not be a feasible assessment format, as it is challenging to standardize the various work products submitted by students and to ensure interrater reliability in scoring student work. A survey, however, can be standardized and norm referenced to allow higher education institutions to make inferences about the ICC of both an individual and a group. Moreover, surveys can include multiple types of selected-response item formats that may better capture the multidimensional nature of ICC. For example, Likert-scale responses may be adequate to capture attitudinal components of ICC, but forced-choice or multiple-choice questions may be more appropriate to assess the knowledge and skills that characterize ICC. In the following section, we discuss the possible item types and their strengths and weaknesses within the category of selected-response items.

Intercultural Competence Selected-Response Item Types

Likert-Scale Items

Most ICC assessments reviewed in this study attempt to capture components of ICC using self-report Likert items. Likert-scale items typically ask the respondents to rate their agreement with a given statement on a scale that ranges from one extreme to another (e.g., *strongly agree* to *strongly disagree*). Some assessments use anchors that directly ask respondents to assess themselves on a particular skill. For example, a behavioral regulation item may ask respondents to indicate whether they would change their behavior in accordance with cultural customs. Another variation across ICC assessments with Likert-scale items is the number of response categories or points on the response scale. Most assessments use a 5-point Likert scale, although others range from a 4-point to a 7-point scale.

Although most of the Likert-type items are self-report, one assessment included in our review used Likert-type responses for peer assessments. The Behavioral Assessment Scale for Intercultural Communication (BASIC; Koester & Olebe, 1989) uses a 4-point Likert scale in a peer rating of intercultural communication effectiveness. This instrument was adapted from Ruben's (1976) behavioral assessment of communication competency for intercultural adaptation. (See Chen, 1992, for a review.) The instrument was designed to fit the context of intercultural roommates in a university setting in which one roommate is native to the United States and the other is an international student. Roommates rate each other on eight items measuring the following aspects of ICC: display of respect, interaction posture, orientation to knowledge, empathy, task-related roles, relational roles, interaction management, and tolerance for ambiguity. Unlike the other ICC assessments, each one-item scale presents the roommate with a behavioral description of the person that they are rating for each of the four points on the Likert scale. The BASIC is the only ICC assessment identified that includes this use of descriptions for Likert-scale anchors (similar to anchored vignettes; G. King, Murray, Salomon, & Tandon, 2004), as the majority of assessments use more traditional Likert-scale response categories (i.e., strongly agree to strongly disagree).

Multiple-Choice Items

To directly measure the knowledge components of ICC (i.e., language and cultural knowledge), multiple-choice items are typically used, such as in the Global Awareness Profile (GAP; Corbitt, 1998) and the Global Competence Aptitude Assessment (W. D. Hunter et al., 2006). These assessments differ in that some multiple-choice items assess cultural knowledge that is general or global and others assess knowledge that is specific to one culture. An example of a global culture item would be something akin to "What is the most popular sport in the world?" As one can see, such an item does not ask about one particular culture, but rather references the general world population.

In addition to culture-general knowledge, the GAP uses multiple-choice items to assess knowledge of the environment, politics, geography, religion, and socioeconomics of six regions (Asia, Africa, North America, South America, the Middle East, and Europe) around the world. In contrast, the Global Competence Aptitude Assessment (Global Leadership

Figure 1 Screenshot of the Implicit Association Test, a test of hidden bias. Retrieved from UnderstandingPrejudice.org (http://understandingprejudice.org). Copyright ©2002–2016 by S. Plous. Reprinted with permission.

Excellence, 2010) uses multiple-choice items based on specific cultures, without any culture-general items. An example of a culture-specific item is, "When greeting a colleague from Chile, one must … " Based on the norms of the culture and context of the situation described, the examinee selects the most appropriate response from a list of choices.

Implicit Association Tests and Q-Sort Methodology

Less common item formats that have been employed to assess the attitudinal component of ICC include implicit association tests (IATs) and the Q-sort methodology. IATs typically capture how strongly a test taker relates two mental representations, or concepts, by measuring the response time (latency) for making the correct association (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). This assumes that the faster a test taker matches an object to a concept, the stronger the relationship is that the test taker perceives between those concepts. One IAT, the Tests of Hidden Bias, assesses negative prejudices toward various ethnic groups by presenting examinees with a photo of a White/Caucasian face next to an African American face on a computer screen and requiring the participant to quickly select the "good" or "bad" photo. Figure 1 presents a screenshot of the free test online. Because 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" (Greenwald et al., 2009, p. 18). Such IATs have been criticized as being too specific to the context of the United States, a country in which race has historically been conceptualized as ethnically dichotomous (i.e., Black vs. White). In response, other IATs have been developed specific to other cultures (e.g., a Romanian IAT; Bazgan & Norel, 2013).

Q-sort is another method that has been used in ICC assessments. The Q-sort methodology has been used in many areas of psychology and involves rank ordering of subjective concepts. The Intercultural Communication and Collaboration Appraisal tool (ICCA) developed by Messner and Schäfer (2012) uses the Q-sort methodology when it requires examinees to sort cards (or concepts, if administered online) in response to a given prompt. The ICCA includes two Q-sorts. The first sort consists of the examinee sorting 48 attitudes, behaviors, and beliefs in order from most descriptive of self to least descriptive. The second sort involves the examinee selecting the most important six intercultural competencies from a set of 12 competencies and ranking them in order of importance.

Situational Judgment Tests

Another method of assessing ICC is the situational judgment test (SJT). SJTs aim to measure an ability or competency based on the participant's choice of response to a hypothetical situation. After reading a few sentences representative of a real-world situation, participants then select the appropriate response option of the presented set or respond to an open-ended prompt. Most of the SJT prompts focus on behavioral and knowledge components. Prompts such as "What would you do?" require the participant to indicate the behavior they would most likely engage in from a series of potential actions (Whetzel & McDaniel, 2009). The options are often scored on a scale of *most effective*, *neutral*, and *ineffective* behavior to produce a composite score for the SJT. Knowledge prompts such as "What is the best answer?" require the

participant to choose the correct answer in the given situation. Sometimes participants are required to rank the responses in order of *most effective* to *least effective* (Whetzel & McDaniel, 2009). According to a recent meta-analysis, SJTs demonstrate substantial criterion, content, and face validity (Whetzel & McDaniel, 2009). For example, McDaniel, Morgeson, Finnegan, Campion, and Braverman's (2001) meta-analysis generated an adjusted correlation of .34 between SJTs and job performance, supporting criterion-related validity of SJTs.

However, due to the multidimensional nature of many SJT items, they typically have low internal consistency as indicated by Cronbach's alpha. Given this reason, experts recommend the use of parallel forms or test–retest reliability when examining the reliability of SJT items instead of using Cronbach's alpha (Whetzel & McDaniel, 2009). The "correct" response option can also be contested, as it is often determined by consensus, which may potentially bias the test. For cross-cultural SJTs, this method may be open to bias if test developers are not conscious of their cultural assumptions. Applicants typically express positivity toward this type of test (Lievens, Peeters, & Schollaert, 2008). Moreover, this test type, by assessing intentions, captures more direct indicators of behavior than attitudinal measures and is well suited to measure skills. Regardless, scores on these items are still not immune to inflation by practice effects and participant deception.

Only a few examples of SJTs exist relevant to ICC context, although 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 cultural scenarios and alternative behavioral options they then discuss (Bhawuk, 2001; Earley & Peterson, 2004). The Cultural Intelligence Assessment (Thomas et al., 2015) asks test takers to choose among a set of behaviors to indicate which one they believe to be the most correct choice for a given scenario. Participants are asked to complete 14 questions designed to measure cultural knowledge, skills, and metacognition. Another SJT, designed to measure cross-cultural social intelligence (CCSI; Ascalon, Schleicher, & Born, 2008), asks participants to rate the likelihood that they would perform each of four behavioral options in response to a series of cross-cultural scenarios. The four options fall into specific categories (nonempathetic, nonethnocentric; nonempathetic, ethnocentric; empathetic, nonethnocentric; and empathetic, ethnocentric), allowing for the creation of two subscales: empathy ($\alpha = .61$) and ethnocentrism ($\alpha = .71$). Coefficient alpha for the overall scale was $\alpha = .68$ (Ascalon et al., 2008).

The CCSI is an example of an SJT measure relevant to ICC that demonstrates evidence of relationships with conceptually related constructs such as cognitive ability (e.g., GMAT; $r = .30$) and personality constructs (Ascalon et al., 2008). The GMAT has been shown to have adequate reliability ($\alpha = .92$ for the test as a whole). Specifically, the relationship between the CCSI scores and three of Goldberg's (1999) International Personality Item Pool (IPIP) subdimensions (conscientiousness, emotional stability, and openness to experience) averaged $r = .30$. The IPIP also demonstrates adequate overall internal reliability ($\alpha = .80$). The CCSI itself has somewhat low reliability ($\alpha = .68$ for the overall, $\alpha = .61$ for the empathy subscale, and $\alpha = .71$ for the ethnocentrism subscale), but these coefficients are roughly similar to other SJT studies (Chan & Schmitt, 1997). Combined, the evidence of internal consistency and convergent validity was taken as a strong indicator of the initial validity of both the measure and the use of SJTs to assess ICC. To the extent of our knowledge, however, no SJT specific to ICC presents evidence of criterion validity (Ascalon et al., 2008).

Simulation-Based Measurement

Although commonly used as training tools for the development of ICC, simulations have also been used to assess ICC (e.g., Harrison, 1992; Jarrell, Alpers, Brown, & Wotring, 2008). Simulations involve role-playing activities in which participants engage in a limited intercultural scenario. The simulation may require the participant to interact with a confederate (a paid assistant who has been instructed to act in a particular way) or an avatar (a figure representing a person or a computer-simulated character) who may be enacting his or her own cultural norms, the cultural norms of a different group, or fictitious norms. Depending on the simulation, other participants in the simulation may play this role instead of confederates. Perhaps the most well-known and commonly conducted intercultural simulation is the BaFa' BaFa' simulation (Shirts, 1977). This simulation requires students to pretend to be in two fictional cultures and interact with each other in order to attempt to collect a certain number of cards, the exact nature of which depends on their culture. The two cultures are loosely designed to polarize individual–collectivism differences (preference for group vs. individual) with verbal and nonverbal differences included (i.e., preference for volume and personal space). Aside from accomplishment of the game goals, observers could also gather interaction data to assess the behavioral component of ICC. This measure would have to be validated, however, as the current simulation kit does not include a behavioral checklist.

A more psychometrically sound example is a simulation by Harrison (1992). This simulation involved participants interacting with a confederate pretending to manage a Japanese employee. The interaction was then independently rated by two judges in terms of maintaining harmony, soliciting employee input, demonstrating personal concern, improving consensus, and reducing conflict (Bhawuk & Brislin, 2000). Another well-known cultural simulator is the Robin Sage Exercise (Skinner, 2002), which serves as the culminating training activity for the Army Special Forces Qualification Course. This 2-week training exercise and assessment involves an intensive military simulation in the fictional country of Pineland, encamping over 8,000 miles of North Carolina and using thousands of volunteers (Parkins & Williams, 2011). Although this exercise has been restricted to the military context, it does expressly assess ICC and therefore demonstrates the use of simulation for ICC measurement.

Validity and Reliability Evidence of Existing Assessments

According to the *Standards for Educational and Psychological Tests* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), every assessment should: (a) produce consistent and accurate scores (reliability) and (b) provide sufficient evidence to support that it accurately measures what it is intended to measure (validity). In this section, we first discuss reliability evidence for the previously developed ICC assessments reviewed in this study. We then discuss the validity evidence regarding the internal structure, the relationships with conceptually related constructs, and the relationship with criteria. A summary of the reliability and validity evidence is presented in Table 3.

Test and Scale Reliability

As previously discussed, the majority of ICC assessments consist exclusively of Likert-type items, and the test and scale reliability evidence was generally adequate. Over 90% of the scales provided evidence of adequate reliability, most commonly assessed via coefficient alpha (α), a measure of the average intercorrelations among test items. However, for ICC assessments with more than one subdomain, several measures with adequate overall alpha values (e.g., Cross-Cultural Adaptability Inventory [CCAI]; Davis & Finney, 2006) had subscale scores that dipped below .70, which is the common cutoff for acceptability (Kline, 2000). Although fewer in number, other scales were able to provide evidence of adequate reliability using test–retest (e.g., Inventory of Cross-Cultural Sensitivity; Bazgan & Norel, 2013) and alternate forms evidence (e.g., Cross-Cultural Sensitivity Scale; Pruegger & Rogers, 1993). For scale-specific reliability information, see Table 3.

Validity Evidence Regarding Internal Structure

One important aspect of validity evidence is the internal structure (i.e., dimensionality) of the assessments, which indicates whether the association among test items corresponds to one or more intended domains (or dimensions) of the assessment (AERA, et al., 2014). One of the most commonly used methods to evaluate the internal structure is confirmatory factor analysis (CFA; Rios & Wells, 2014). An acceptable index of model fit indicates that the structure of the assessment is as intended, based on the relationship between the test items and the construct(s).

Among all the ICC assessments in Table 3, more than 10 assessments reported a single overall score to test takers, and five of them provided evidence to support the unidimensional structure of the assessment. Graf and Mertesacker (2009) fitted a one-factor model to data from the Nonverbal Communication Competence Scale, and the results suggested that all items were measuring the same construct. Arasaratnam (2009) and Olebe and Koester (1989) also provided similar evidence for the Intercultural Communication Competence test and the BASIC test, respectively.

For assessments that report subscale scores, about half provided evidence to support the multidimensional structure of the assessment. For example, the CFA results from Wang et al. (2003) suggested the four subscales of the Scale of Ethnocultural Empathy were adequately measuring the intended constructs, and the four factors shared approximately 81% of the total variance. Hammer et al. (2003) also reported a good model fit of a five-factor model for the Intercultural Development Inventory. However, a multidimensional structure of assessments is not always supported by the data. For instance, Davis and Finney (2006) found weak support for the four-factor model originally proposed for the CCAI. Nguyen, Biderman, and McNary (2010) also found each item from the CCAI loaded on a general factor (i.e., cross-cultural

Table 3 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 responsibility; alphas ranging from .657 (cognitive – knowing) to .773 (cognitive – knowledge)	Relationship with other assessments: t tests revealed that initial scores on the Global 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 a single 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.

Table 3 Continued

Test	Reliability	Validity
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 0.16) with the Intercultural Anxiety Scale (Stephan & Stephan, 1985). 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 (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 ($r = .23$, $p < 0.05$).

Table 3 Continued

Test	Reliability	Validity
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 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.

Table 3 Continued

Test	Reliability	Validity
Tests for hidden bias Miville-Guzman Universality–Diversity Scale (M–GUDS)	Not significantly different from other IATs Alphas range between .89 and .94.	Relationship with other assessments: not significantly different from other IATs. 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. Subscale: alpha = .37–.73 for Canadian sample and .25–.55 for Japanese sample.	Internal structure: a moderate fit five-factor solution from both the 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.

Table 4 Continued

Test	Reliability	Validity
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)		Internal structure: There are two factors: collectivism and individualism.
Nonverbal Communication Competence Scale (NVCCS)	Alpha for College of Business sample: .82; Alpha for East–West Center sample: .84. 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 ($r = .219$).
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).

adaptability) and one of the nine group factors (e.g., emotional resilience, flexibility/openness, personal autonomy, and the like). These group factors represented the constructs that were not accounted for by the general factor. Therefore, even though the CCAI reported four subscale scores, the results from the two studies did not support a four-dimensional structure of the assessment. In sum, evidence supporting the multidimensional structure for existing ICC measures is not as strong as desired.

Further, about half of the ICC assessments reviewed in this paper did not report evidence of adequate internal structure. Best practices for scale construction support providing this evidence by demonstrating good model fit of an item-level factor analysis. Best practices for scale construction suggest that this evidence is ideally provided by demonstrating good model fit of an item-level factor analysis. For example, the Global Competencies Inventory (GCI; Bird, Stevens, Mendenhall, & Oddou, 2002) reported only the correlation among the three subscores instead of the measure's internal structure. The lack of evidence describing the structure of the scale demonstrated a significant gap in validity evidence and thus a particularly notable weakness.

Validity Evidence Regarding Relationships With Conceptually Related Constructs

The second aspect of validity evidence is the relationship with conceptually related constructs, traditionally known as convergent and discriminant validity. A correlation coefficient between two assessments is typically used to estimate the degree to which the constructs measured by the two assessments are related to each other. According to *Standards* (AERA, et al., 2014), a valid assessment would show correspondence with relevant constructs and discrimination with irrelevant constructs. Because the correlation coefficient is affected 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 important to report the reliability information along with the correlation coefficient. Overall, about half of the existing ICC assessments reviewed in this study provided some evidence concerning a relationship with related constructs.

Research with the popular cultural intelligence construct has fairly ample evidence, primarily from organizational samples (Leung et al., 2014), but also in educational contexts. For example, Erez and colleagues (Erez et al., 2013; Lisak & Erez, 2015) conducted two studies using the Cultural Intelligence Scale (Ang, Van Dyne, & Koh, 2006; Ang et al., 2007) with students participating in cross-cultural virtual team projects. The results demonstrated a strong relationship ($r = .50$) between the cultural intelligence of students in global virtual teams and a sense of belonging to global context, termed *global identities* (Erez & Gati, 2004). The researchers measured global identities with a validated and adequately reliable Global Identity Scale ($\alpha = .85$; Erez & Gati, 2004; Shokef & Erez, 2006, 2008). One of the studies further connected cultural intelligence to openness to cultural diversity ($r = .16$) and leadership emergence ($r = .56$; Lisak & Erez, 2015). Providing some evidence of an antecedent in the nomological network of ICC, other research with this scale connected it to expectancy disconfirmation after cooperative intercultural contact (Rosenblatt et al., 2013).

In a study by Hammer et al. (2003), the authors confirmed the theoretically postulated relationships among the subscales of the Intercultural Development Inventory (IDI; $\alpha = .80 – .85$) and two related assessments—the Worldmindedness Scale ($\alpha = .67$) and the Intercultural Anxiety Scale ($\alpha = .86$). Higher scores on the denial/defense subscale of the IDI were related to lower scores on the Worldmindedness Scale ($r = -.29$) and higher scores on the Intercultural Anxiety Scale ($r = .16$).

Structural equation modeling, which models error terms in order to isolate the latent construct, constitutes another, more robust, method of supporting relationships among measures. Instead of calculating the correlation coefficient from observed scores, Nguyen et al. (2010) used a structural equation modeling technique to examine the relationship between the CCAI and Goldberg's IPIP Big Five questionnaire (Goldberg, 1999). The results showed weak to moderate correlations between the two assessments ($r = .18 – .55$), which suggests that test takers with better cross-cultural adaptability tend to be more extroverted, agreeable, conscientious, emotionally stable, and open to new experiences. The correlation coefficient estimated from the structural equation model is the correlation between the underlying constructs of two assessments. Unlike the statistics employed in the Hammer et al. (2003) study, measurement error does not affect the structural equation model correlations. Therefore, structural equation modeling is a promising method for future research to provide validity information regarding relationships with conceptually related constructs.

Validity Evidence Regarding Relationship With Criteria

The relationship between the assessment and related criterion measures is another important aspect of validity evidence (AERA et al., 2014). Examples of the criteria used for existing ICC assessments include self-evaluation, peer impressions, job performance, and the like. Few of the assessments in Table 3 provide this type of validity evidence, perhaps due to the resource-heavy requirements of criterion data collection.

Nguyen et al. (2010) examined whether the subscale scores of the CCAI would predict the number of international job assignments when controlling for the variance of the general factor (cross-cultural adaptability). The results partially supported the hypothesis, as only two subscales (resilience and personal autonomy) were weakly correlated with the logarithm number of international job assignments ($r = .20$ and $r = .29$, respectively), and no subscales were correlated with the actual number of assignments. In a study by Matsumoto et al. (2001), the participants who took the Intercultural Adjustment Potential Scale (ICAPS) also rated themselves and all other members of the focus group on a two-item rating scale about intercultural adjustment. Two interviewers also made both ratings of all participants. The analysis showed the composite score of the ICAPS was significantly correlated with self, peer, and interviewer ratings ($r = .69, .70,$ and $.66,$ respectively; $p < .001$), which supported the utility of the ICAPS in predicting intercultural adjustment. In addition, the Miville-Guzman Universality–Diversity Scale, which measures awareness and potential acceptance of both similarities and differences in others, was not significantly related to the $SAT^{®}$ verbal scores (Miville et al., 1999), providing evidence of discriminant construct validity. However, in a U.K.-based study of students in culturally diverse teams, the Multicultural Personality Questionnaire was found to be related to exam grades (Van der Zee, Atsma, & Brodbeck, 2004); in particular, the flexibility component was moderately related using hierarchical linear modeling ($z = 1.78$).

In a study with 71 recruiters in a U.S. high-tech organization (Hammer, 2011), scores on the IDI were found to be correlated ($r = .43$) with the rating of success in meeting diversity goals for recruitment. In another funded study on study abroad students (Hammer, 2005), 1,500 students completing a 10-month homestay program organized by AFS Intercultural Programs, an American-based study abroad facilitator, were compared to a control group ($n = 638$) of students who remained at their home institutions. Students involved in the homestay program resided in Austria, Brazil, Costa Rica, Ecuador, Germany, Hong Kong, Italy, Japan, and the United States. Scores on the IDI were found to be positively correlated with the number of intercultural friends students reported having, a sociometric measure of experience success reflecting the ability of students to build international relational networks (Hammer, 2005). The measure was also found to be related to reduced anxiety and increased satisfaction with the experience.

Other evidence suggested that the Cultural Intelligence Scale (CQS) may relate to several valued student outcomes. In particular, higher scores on the CQS were related to commitment to and satisfaction with international educational courses (e.g., Morell, Ravlin, Ramsey, & Ward, 2013; Ramsey, Barakat, & Aad, 2014), intention to work abroad (e.g., Remhof, Gunkel, & Schlaegel, 2013), and global virtual team leadership (Erez et al., 2013; Lisak & Erez, 2015). These outcomes, which fall into the category often labeled *previous experience*, serve as useful criteria as they have been related to global leadership effectiveness (e.g., Caligiuri & Tarique, 2012). Research also suggests that study abroad experiences develop student competencies when assessed using this scale (Engle & Crowne, 2014; Varela & Gatlin-Watts, 2013). However, the validity evidence relating the scale with adjustment while studying abroad is mixed. One study, with international students studying in New Zealand, indicated that the motivational subscale was not predictive of psychological adjustment during study abroad (Ward, Wilson, & Fischer, 2011); another study, with a Taiwanese sample, indicated that cultural intelligence was not related to adjustment (Lin, Chen, & Song, 2012). It should be noted that the two studies used different scales for adjustment—the Sociocultural Adaptation Scale (Ward & Kennedy, 1999) and the Black and Stephens (1989) scale measuring work, interactional, and general adjustment. The Black and Stephens scale is commonly used, but has several measurement concerns, including proper validation evidence (Thomas & Lazarova, 2006).

Summary of Reliability and Validity Evidence

The review of the reliability evidence of existing ICC assessments suggests no major issues with reliability at the total test level. All the assessments in Table 3 reported reliability evidence suggesting satisfactory reliability at the test level; however, some minor issues still exist. One issue is that the subscale score reliability of five assessments was found to be unsatisfactory ($\alpha < .70$), including the Global Perspectives Inventory, Cultural Intelligence Assessment, and CCAI. As subscale scores are usually reported for diagnostic purposes (e.g., when used as a training tool), unreliable subscores may

result in inaccurate diagnoses and, therefore, provide misleading information for score users. Unreliable subscales suggest that error will contaminate different facets unequally and reduce the quality of a development plan constructed based on scores. Further, it would be difficult to validate ICC training interventions when some subscale scores randomly fluctuate. Another issue observed is related to the comparability among test forms. Of the three ICC assessments in Table 3 that consisted of more than one test form, two reported high correlations between test forms, although one did not provide any information.

Unlike the reliability evidence, the quantity and quality of validity evidence varied significantly among existing ICC assessments. Roughly half of the assessments in Table 3 reported validity evidence regarding internal structure, about half reported evidence regarding the relationship with related constructs, less than one third reported evidence regarding the relationship with related criteria, and only two assessments reported all three aspects of validity evidence. In addition to quantity, the quality of some available validity evidence was also unsatisfactory. For instance, the hypothesized internal structure of some assessments was not supported by the data, which raises questions about subscale score reporting. The relation between some ICC assessments and their related measures were also poorly estimated due to the low reliability of the tests.

In general, stronger validity evidence was available for some assessments developed after 2000 (e.g., the Cultural Intelligence Scale and the IDI) and the assessments developed by organizations (e.g., the CCAI). However, for most assessments developed 20 or 30 years ago or developed by independent researchers, relatively insufficient validity evidence exists. This lack of validity evidence may be attributable to limitations on resources such as financial support or available statistical packages, but may also reflect an outdated approach to validity. After Messick (1995) described validity as a single construct for which researchers could provide various types of evidence, the importance of gathering a range of validity evidence to support test score inferences has been gradually acknowledged by test developers. Although more validity research has been conducted in recent years, one aspect of validity that is still often missing is the evidence regarding the relationship with criteria. This holdover may explain the prevalence of validity evidence limited to a single type. In keeping with Messick, no priority was given to any type of evidence; however, the particular lack of criteria-related evidence should be highlighted. Very few measures were related to any sort of accepted criteria. Therefore, future validity research should be encouraged to gather criteria information to clarify the extent to which the scores from an ICC assessment predict test takers' skills to communicate and work across cultures in authentic situations. Criteria-related evidence is particularly convincing in terms of investment—if a strong argument is to be built for higher education to invest in the development of these skills, then persuasive evidence of their relations to valued outcomes will be the best foundation.

Challenges in Designing an Intercultural Competence Assessment

Confounds and Issues With Self-Report Measures

Self-report measures are a versatile tool suited for capturing attitudes and declarative knowledge (Gabrenya, Griffith, Moukarzel, Pomerance, & Reid, 2012). For the assessment of ICC, however, sole reliance on self-report measures presents several challenges. First, it may be confounded with student experience levels. The typical young adult will have limited exposure to multicultural environments and less experience reflecting upon the skills and behaviors comprised by ICC. Thus, items that rely on previous experience may be adversely impacted by the lack of exposure. Other confounds include cognitive biases, in particular future-oriented optimism (e.g., Bazerman, 1990), which may further complicate self-report as students respond to items based on their most idealistic self. Additionally, self-report items may be inappropriate for assessing interaction tendencies and other ICC skill components.

Moreover,## although the current self-report assessments seem to reliably measure the attitudinal components of ICC, faking behaviors may present an additional challenge for self-report measures (Likert-scale responses). The tendency for respondents to deliberately provide inaccurate responses or self-descriptions to make themselves appear more attractive, interesting, or valuable (faking) is a critical concern in self-report attitudinal measures such as those on ICC assessments. As previous research has demonstrated a large impact of faking on test results ($d = 0.48$ to $d = 3.34$; Viswesvaran & Ones, 1999), researchers have attempted to control for it by (a) identifying and making statistical adjustments and (b) developing item types that make it more difficult for respondents to fake.

Faking

Self-report respondents can engage in faking behaviors intentionally and unintentionally. For many years, faking behavior was conceptualized as socially desirable responding. Seminal work by Paulhus (1984) suggested that social desirability comprises two components: self-deceptive enhancement (SDE) and impression management (IM). SDE was considered an unconscious form of social desirability that is associated with a positive outlook (Taylor & Brown, 1988). IM, on the other hand, is an intentional attempt at deception (Paulhus, 1984). It is likely that this two-factor structure of social desirability was implicitly extended to faking behavior because of the literature's close association of the two phenomena. More recent faking research now makes a distinction between unintentional misrepresentation, which is akin to bias, and intentional applicant faking behavior (e.g., McFarland & Ryan, 2006; Sackett, 2011). In the case of SDE, the source of bias is a general tendency to have positive views of oneself (Taylor & Brown, 1988). Other biases may also contribute to inflated scores under motivated conditions. For example, the future orientation cognitive bias influences respondents to respond more positively to items in the future than the past (Taylor, 1989). Extreme response styles (e.g., using only the ends of a Likert scale) can also distort self-report data (Johnson, Shavitt, & Holbrook, 2011). Even if committed unintentionally, faking behavior still represents a minor threat to validity due to the introduction of additional error variance. This error variance is not likely to be uniform across all respondents, so the impact of unintentional distortion bias is likely small decrements to validity due to the introduction of variance not associated with the target construct. However, practically significant drops in validity are not likely. Owing to this shift in the conceptualization of faking behavior and the low severity of the psychometric consequences, most attention is now focused on intentional faking (Ziegler, MacCann, & Roberts, 2011).

Significant differences in responses across motivated and unmotivated conditions have provided evidence for intentional faking behavior. R. L. Griffith, Chmielowski, and Yoshita (2007) investigated within-person differences in faking behavior across settings. They asked participants to complete a measure of conscientiousness as part of an actual employment application process. Afterward, the researchers contacted the participants and instructed them to complete the same measure as honestly as possible with the reassurance that the second version was for research purposes only. The researchers found a significant difference between responses across the two conditions: Significant within-person differences existed between mean level scores in the applicant condition and mean level scores in the honest condition, $F(2, 59) = 42.32$, $p < 0.001$, suggesting that people can and do intentionally alter their responses in an effort to portray themselves in a more positive light when motivated to do so (R. L. Griffith & Peterson, 2008). This finding suggests that, depending on the environment, test takers are not always honest or accurate or both on self-report tests. The pattern of within-subject score inflation has been replicated when data was collected in the same fashion (e.g. Arthur, Glaze, Villado, & Taylor, 2010; Peterson, Griffith, Isaacson, O'Connell, & Mangos, 2011). R. L. Griffith and Converse (2011) synthesized the empirical literature via statistical analyses, simulations, and logical deduction and estimated that, on average, 30% of applicants ($\pm 10\%$) engage in faking behavior. The impact of faking behavior is substantial, with decrements on internal (Chaney & Christiansen, 2004) and external validity metrics (e.g., Komar, Brown, Komar, & Robie, 2008; Peterson et al., 2011). Some of the decrement to validity may be artifactual as a result of nonlinearity in the data (Peterson & Griffith, 2006). Applicants who increase their scores, but perform at a level predicted by their true score, provide data points that function as outliers. Essentially, the faker's data points are shifted toward the higher end of the personality score distribution, but their performance is not commensurate with this positive shift in scores. This deviation from the monotonic relationship between personality and performance results in a nonlinear artifact that attenuates the correlation between the personality measure and the outcomes of interest (Peterson & Griffith, 2006). Other contributing factors to the attenuation of predictor criterion relationships may be more substantive in nature. Some research has demonstrated a significant relationship with applicant faking and counterproductive behaviors in the workplace (Peterson et al., 2011).

Administering External Items

One approach to controlling for faking consists of administering external items that are unrelated to the construct of interest (e.g., ICC) and do not count toward the examinee's score. Currently, there are two types of external items: (a) bogus and (b) social desirability items. Bogus external items are ones that appear to be related to the construct (e.g., ICC), trait, skill, or task of interest, but the objects or scenarios described in the items do not actually exist (e.g., "How often do you utilize murray-web system to locate unpublished research articles?"; where the murray-web system does not exist; Dwight & Donovan, 2003, p. 10). In contrast, social desirability items measure the tendency to answer questions

Directions: Out of the three statements, select one that describes you MOST
accurately and one that describes you LEAST accurately.

	MOST like me	LEAST like me
I am relaxed most of the time		
I start conversations		
I catch on to things quickly		

Figure 2 A forced-choice item asks the respondent to choose from one of two or more options that appear equally desirable.

in a manner that is perceived to be viewed favorably by others. Consistent endorsement of either item type may suggest that respondents are providing unauthentic or faked responses. Even though social desirability items are often used as proxies for faking behavior, research has suggested that they are ineffective at identifying and controlling for faking (R. L. Griffith & Peterson, 2008). This research analyzed the validity of social desirability as a proxy for within-subject score change across motivated and unmotivated conditions. Using the proxy variable estimation suggested by J. E. Hunter and Schmidt (2004), R. L. Griffith and Peterson (2008) reported that the operational quality of a measure of social desirability as a proxy for faking was poor (interpreted similarly to a corrected correlation coefficient, between .08 and .11). J. E. Hunter and Schmidt proposed that the quality of a proxy variable could be determined by multiplying the reliability of the proxy measure by the correlation of the proxy measure and the variable of interest. Measures of social desirability are often self-report and demonstrate adequate reliability; however, the correlations between measures of social desirability and within-subject score change are quite low and, in some instances, negative (R. L. Griffith, Malm, English, Yoshita, & Gujar, 2006). Thus, the low proxy index reported by R. L. Griffith and Peterson was influenced more by the lack of common variance of measures of social desirability than it was by error variance. In general, social desirability items are no longer viewed as a useful tool to assess and correct for faking behavior.

When using external items, two approaches are available to control for the impact of faking on test scores: (a) deletion of the data from respondents deemed to be faking and (b) statistical adjustments. The first approach is the older of the two and consists of setting an a priori threshold for the number or percentage of bogus or social desirability items endorsed. If examinees exceed this a priori threshold, they are deemed to be faking, and their data on the assessment of interest is completely deleted. The second approach is to compute corrected scores for respondents who provide unauthentic responses by regressing social desirability scores onto trait scores (e.g., ICC) to compute a residual score. This approach attempts to parcel out variance associated with social desirability from the construct of interest (ICC); however, research has shown that this partialing may remove meaningful variance, which leads to a decrease in the validity of the measure (e.g., Soubelet & Salthouse, 2011).

Employing Alternative Item Types

As the use of external items merely attempts to identify faking behavior, researchers have attempted to apply alternative item types (i.e., non-Likert items) to make it more difficult for examinees to fake. Such an approach does not purport to completely eliminate faking and still involves the use of self-report, but it does aim to reduce it. For this purpose, two item types have been proposed: (a) SJT and (b) forced-choice items. As described previously, SJTs present a respondent with a task-related situation, which can be in written, video-based, or multimedia format, and they ask the respondent how she or he would theoretically respond (i.e., not based on actual behavior) by choosing from a list of options (Whetzel & McDaniel, 2009).

In contrast, forced-choice items ask the respondent to choose from one of two or more options that appear equally desirable (Christiansen, Burns, & Montgomery, 2005). As an example, Brown and Maydeu-Olivares (2011) developed a forced-choice triad item for a Big Five personality inventory (see Figure 2).

Although both SJTs and forced-choice items have been proposed as item types that can reduce faking, more research has been conducted on the latter item type. Specifically, when comparing Likert and forced-choice items, the latter have been shown to significantly reduce the impact of faking on mean scores by as much as 0.68 standard deviations (Jackson, Wroblewski, & Ashton, 2000; Martin, Bowen, & Hunt, 2002). However, forced-choice items provide two limitations when compared to Likert items: (a) They require an increased number of items and (b) there are a number of psychometric concerns related to scoring. Regrettably, very little research has investigated whether using forced-choice items is worthwhile

in low-stakes testing contexts, as there is uncertainty regarding the impact of faking in such a context. Assuming that faking is an issue on the ICC assessment, the best approach may be to use multiple item types, particularly as forced-choice items will require increased test length.

Culture-Specific Versus Culture-General Knowledge

A known challenge to assessing the knowledge and skills associated with ICC is that they can be context dependent. For example, cultural knowledge is often situated within a specific culture and may require specific language skills. However, assessing ICC with items referencing a specific culture may be unfeasible: An individual may come into contact with a number of different cultures within his or her lifetime. As a result, it may be preferable to assess culture-general knowledge or knowledge that is useful in interpreting, coping with, and adapting to cross-cultural interactions. That is, instead of assessing how knowledgeable an individual is about the cultural norms and practices of a particular country or region, the more desirable approach may be to assess an individual's recognition that a new situation may be influenced by cultural differences. This recognition is largely developed through a cultural schema, which is a mental structure, framework, or system that is used to understand how personal background, values, and beliefs impact cross-cultural interactions (Brenneman et al., 2016). This culture-general position has also gained ground in the cross-cultural training literature (e.g., Brandl & Neyer, 2009). Thus, scenario-based items may be more appropriate than self-reported items, which is an issue discussed in the next section.

Capturing the Interactional Component of Intercultural Competence

One of the challenges of assessing ICC is that the construct is composed of attitude, knowledge, and skill subdomains that require an interpersonal interaction to occur in order to be assessed. As an example, an individual may have to realize that he or she is in a situation where cultural differences may be influential, hypothesize how the situation is going to unfold, decide how to behave, and take a course of action (Brenneman et al., 2016). Such an interaction is dynamic in nature and must be simulated through a scenario. However, building such scenarios requires a heavy expenditure of resources, complete with high development costs and overhead. The aforementioned BaFa' BaFa' takes about 2 hours for 20 people to complete, making it a logistical challenge to administer with even the smallest collegiate population. Although video- or avatar-based simulations represent one exciting potential alternative to in-person simulations, they, too, require a substantial investment of time and money. An additional option could be to use SJTs. This method of assessment has been attempted in the Cultural Intelligence Assessment (Thomas et al., 2015), but limited validation evidence prevents firm inference on the use of this technique. Moreover, some scholars argue that even a simulated scenario fails to mimic the dynamic nature in which ICC is negotiated between two or more parties. In sum, assessing the real-world dynamic of ICC is a great challenge that requires creativity, particularly when considering practical constraints, although some recent projects are making strong inroads using virtual platforms.

Inadequate Predictive Validity

Because ICC is a complex skill, it is sometimes difficult to find an appropriate criterion to evaluate the predictive validity of an ICC assessment. As previously discussed, the existing ICC assessments were developed for various purposes; thus, the choice of criterion in current validity research varies considerably. The variability of criteria raises a concern regarding the reliability of the criterion measures, given that a poor measure of the criterion may hinder validity evidence. Therefore, one challenge is to determine the definition of ICC in higher education and identify acceptable and reliable criteria measures to establish predictive validity evidence. One purpose of measuring college students' ICC as one of their learning outcomes is to predict if they are able to effectively communicate and work in an organization with global missions. At this point, however, it is unclear if such organizations would provide information about their current employees' communication capacity and work efficiency in order to establish evidence of predictive validity. Therefore, given these challenges, obtaining criterion measures will be an ongoing process and one that may require longitudinal research to establish predictive validity evidence for ICC assessments in higher education.

Summary

These measurement concerns (respondent faking, adequate predictive validity, and incorporation of the interactional and culture-general domain without overreliance on specific culture content) challenge those seeking to assess ICC. Furthermore, conceptual concerns regarding existing ICC models also complicate the task. A useful framework for ICC must provide specific definitions, clearly delineate between the construct and its nomological network, incorporate both the cognitive and noncognitive subdimensions, and clarify the relationships between the subdimensions. Moreover, such a framework offers the most utility when constructed to redress the measurement concerns described herein. Based on all the above reasons, a new framework designed to overcome both sets of concerns is developed.

A Proposed Framework for Intercultural Competence in Higher Education

Operational Definition of Intercultural Competence

Synthesizing the models from which the reviewed scales were created (e.g., Ang et al., 2007) as well as empirical research (e.g., Abbe, Gulick, & Herman, 2007), we propose a framework and operational definition to serve as the basis for the development of a new assessment of ICC (Table 4). We propose a new framework here for several reasons. First, many existing frameworks do not offer insights on how to translate the theoretical definitions into actual assessments, which may have contributed to the difficulty in accumulating validity evidence. The proposed framework aims to provide an elaborated discussion of assessment considerations that may better guide the development of an operational assessment. Second, academic experts on ICC remain divided, such that many existing models have no widespread support outside of their own particular camp of researchers. This tendency is apparent in the trend for ICC validity evidence to be collected primarily by those whose names are attached to the development of the assessment (e.g., Ang et al., 2007). Third, developing a new model provides the opportunity to tailor it to the purpose of the assessment and its target population (i.e., higher education), focusing on developable skills and excluding components that are less directly related to successful achievement of intercultural goals. More important, generating a new model creates the opportunity to address the various concerns regarding construct validity discussed in the previous sections. For example, we theorize that the ability to acquire declarative cultural knowledge is less predictive of success than the ability to apply relevant cultural knowledge during an intercultural interaction. Thus, we propose the following framework.

To begin, we draw on a definition from prior research: ICC "reflects a person's capability to gather, interpret, and act upon these radically different cues to function effectively across cultural settings or in a multicultural situation" (Earley & Peterson, 2004, p. 105). Next, we propose a framework that builds on a process model of social thinking (Grossman, Thayer, Shuffler, Burke, & Salas, 2015) by splitting cross-cultural interactions into three stages and specifying the skills necessary to support successful performance in each stage. This process model breaks individual behavior in a complex social situation down into four stages (scan, appraise, interpret, and interact) and the cognitive and behavioral skills that support them. In this way, the ICC framework is also developed. Intercultural interaction may be conceptualized as occurring in three stages: approach, analyze, and act (see Figure 3). These stages act as the dimensions of the framework. The approach dimension includes the characteristics that impact the likelihood that an individual will initiate and maintain intercultural contact voluntarily, as well as those traits that will define the overall positivity with which an individual responds to cross-cultural interactions. These characteristics include a positive cultural orientation, a tolerance for ambiguity, and self-efficacy. The analyze dimension captures an individual's ability to take in, evaluate, and synthesize relevant information without the bias of preconceived judgments and stereotyped thinking. The analyze dimension includes the following traits: self-awareness, social monitoring, perspective taking/suspending judgment, and cultural knowledge application. The act dimension incorporates the behaviors determined by the previous dimension to assess individuals' ability to translate thought into action while maintaining control in potentially challenging and stressful situations. The act dimension includes behavioral regulation and emotional regulation. The following sections provide more detail about the nature of each trait and skill. Operational definitions can be found in Table 4.

Approach

As specified above, this dimension includes a positive cultural orientation, tolerance for ambiguity, and cultural self-efficacy. Although similar to a general positive attitude toward intercultural situations, a positive cultural orientation is

Table 4 Assessment Framework for Intercultural Competence

Construct domain	Definition	Examples of assessment topics
Approach		
Positive cultural orientation	Evaluation of cross-cultural situations as favorable	• Curiosity and respect for other cultures • Open-mindedness toward cultural differences
Tolerance for ambiguity	Ability to maintain composure and well-being in uncertain or volatile situations without compromising effectiveness	• Tendency to remain engaged in and seek out intercultural interactions despite the innate uncertainty and unpredictability
Cross-cultural self-efficacy	Degree to which an individual believes that he or she can achieve a goal	• Initiation and development of strong rapport with culturally different others
Analyze		
Self-awareness	Degree to which an individual understands the impact of his or her own culture, values, preferences, and previous experiences on his or her cognitive, emotional, and behavioral responses	• Consideration of self as over an individual and a product of his or her culture • Ability to dissect one's own worldview to identify and distinguish between influences of personal history versus influences of culture • Understanding that individuals from other cultures have different worldviews
Social monitoring	Attention to the other's physical, verbal, and nonverbal behaviors and cues during a social interaction; attention to other's responses to one's own actions and signals	• Ability to infer social norms, hierarchies, and interpersonal relationship networks
Perspective taking/suspending judgment	Active consideration of others' potential viewpoints / active refrainment from preconceived cultural schema interfering with information processing	• Removal of one's own stereotyped or heuristic thinking; replace with effortful cognitions regarding other person's viewpoint, motivation, and assumptions • Low reliance on one's own cultural schemas to understand another's culturally different viewpoint
Cultural knowledge application	Utilization of relevant declarative cultural knowledge in an interaction	• Integration of culture-general, culture-specific, historical, and geopolitical information • Actively seeks and uses cultural information in evaluation and decision-making processes
Act		
Behavior regulation	Active monitoring and revision of personal behavior to engage in culturally appropriate behavior and avoid engaging in culturally inappropriate behavior	• Suppression of familiar behaviors when culturally inappropriate; generation of appropriate behaviors
Emotion regulation	The ability to monitor and revise emotions in an automatic or controlled manner	• Control over which emotions are experienced, how and when they are experienced, and how and when they are expressed
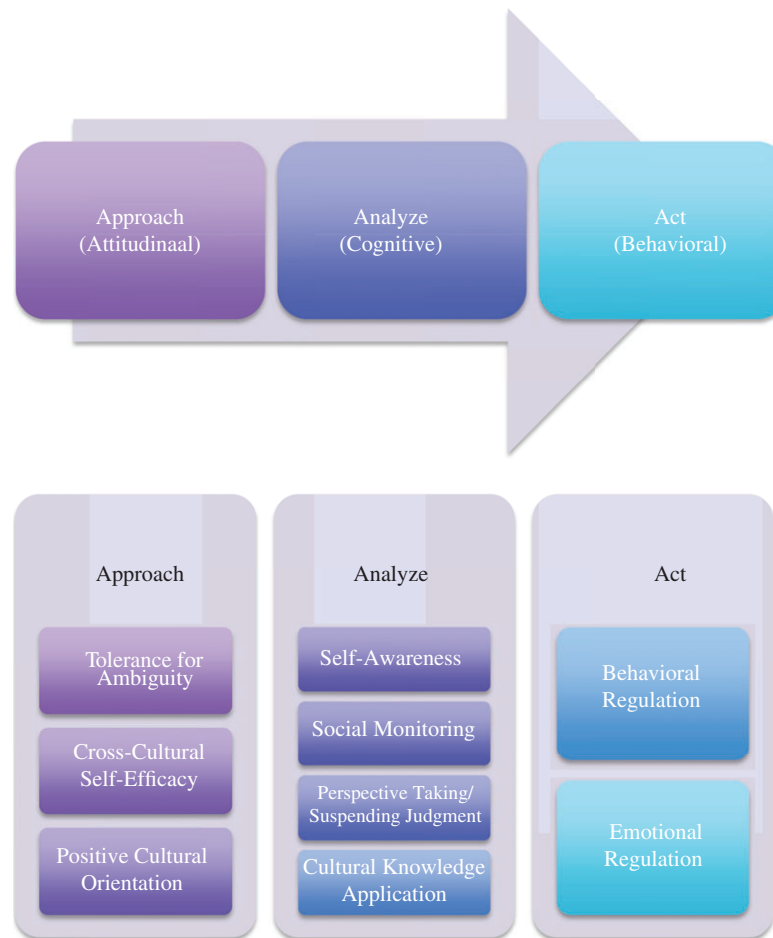
Figure 3 The conceptual model of the *approach*, *analyze*, and *act* intercultural competence framework.

a consolidated representation of several related concepts in the literature. These concepts include cosmopolitanism (i.e., reduced ethnocentrism; Beechler & Javidan, 2007; Levy, Beechler, Taylor, & Boyacigiller, 2007), open-mindedness (Terrell & Rosenbusch, 2013), inquisitiveness (Black, Mobley, & Weldon, 2005), as well as curiosity and respect for other cultures (Beechler & Javidan, 2007). Evidence also suggests that such orientations or attitudes can be changed (Ajzen, 2001). For example, global leadership development programs have been found to foster open-mindedness through participants' genuine curiosity and an attitude of discovery and exploration (Terrell & Rosenbusch, 2013). Therefore, it is possible to conclude that positive cultural orientation is not only malleable but could also predict competencies similar to ICC, such as intercultural sensitivity and global leadership effectiveness (Cushner, 1986; Terrell & Rosenbusch, 2013).

The second subdimension of approach, a tolerance for ambiguity, is repeatedly identified as essential to ICC due to the inherent nature of interacting with individuals from different cultural backgrounds (e.g., Caligiuri & Tarique, 2012). Differences in behaviors, assumptions, communication, and the resulting inability to anticipate potential situations all contribute to the ambiguous nature of intercultural interactions (Lane, Maznevski, & Mendenhall, 2004). Individuals who can tolerate ambiguity not only function effectively in spite of stress (Caligiuri & Tarique, 2012), but also will be less negatively impacted by the stress of the intercultural interaction and more likely to remain engaged and even seek out these situations. Therefore, due to the inherent uncertainty associated with cross-cultural interactions, a tolerance for ambiguity is an important subdimension of the first dimension in ICC.

Cultural self-efficacy is the last subdimension of approach. Self-efficacy influences the challenges in which an individual chooses to engage and his or her attitude toward those challenges. For example, an individual with high self-efficacy in intercultural situations believes that he or she can develop a strong rapport with someone from another culture. Because of this perception, the individual is more likely to initiate and engage in interactions that require development of rapport

with culturally different others. In this way, an individual's level of ICC in part depends on the individual's evaluation of his or her own abilities.

Analyze

This dimension includes self-awareness, social monitoring, suspending judgment, perspective taking, and cultural knowledge application. Self-awareness requires individuals to consider themselves as both an individual and as a member of their own culture. Highly self-aware individuals are capable of dissecting their worldview to identify the influences of their personal history as separate from the influences of their culture, and they understand that different backgrounds will have different worldviews (Reid, Kaloydis, Sudduth, & Greene-Sands, 2012).

Social monitoring includes the ability to infer social norms, hierarchies, and interpersonal relationship networks (e.g., Lodder, Scholte, Goossens, Engels, & Verhagen, 2016). Evidence from neuropsychology suggests that we use social cues, such as expressions, as information to evaluate our performance (Boksem, Ruys, & Aarts, 2011). In the absence of familiar norms, then, social monitoring can provide necessary information to supplement missing native knowledge and evaluate the success of one's chosen course of action, making it a necessary skill for engaging in novel cross-cultural situations.

Suspending judgment and perspective taking are two complementary skills that involve processing situational information without strong personal bias. An individual who suspends judgment removes his or her stereotyped or heuristic thinking; perspective taking replaces these thought patterns with effortful cognitions regarding the other person's viewpoint, motivation, and assumptions. In doing so, individuals reduce their reliance on their own cultural schema in order to act on their understanding of a cultural other's viewpoint.

Cultural knowledge application requires individuals to consider a broad range of information including culture-general information (e.g., cultural value dimensions; Hofstede, 1980), culture-specific information (e.g., French greetings), and historical as well as geopolitical information (e.g., the trends of power and privilege; Hammer, 2012). This skill explicitly refers to the ability of individuals to actively seek and use cultural information in their evaluation and decision-making processes.

Act

This dimension includes behavior regulation and emotion regulation. Behavior regulation is essential to ICC because behavior patterns considered normal in one culture may be inappropriate in cross-cultural situations. Individuals skilled at behavior regulation would be able to suppress any familiar behaviors inappropriate to the cultural context, generate the appropriate behavior for that situation, or perhaps choose not to engage in any behavior at all (e.g., Ang et al., 2007).

Emotion regulation allows individuals to control which emotions they experience, how and when they experience them, and how and when they are expressed (Gross, Salovey, Rosenberg, & Fredrickson, 1998). Because cross-cultural experiences are inherently emotional (e.g., Haslberger, Brewster, & Hippler, 2013; Shaffer, 2012), evidence has suggested that individuals with strong emotion regulation abilities can act more effectively in cross-cultural situations than those without emotion regulation abilities (Haslberger et al., 2013).

The current framework aims to address the particular construct validity challenges of ICC and the criteria highlighted in previous sections (see Validity Evidence Regarding Relationships With Conceptually Related Constructs) First, this framework is grounded in a definition of ICC that offers more clarity and distinguishes it from similar constructs, such as global leadership. Second, the framework demonstrates comprehensiveness; each subdimension assessment includes skills encompassed in other frameworks (e.g., Reid et al., 2012). The framework also expands the comprehensiveness of ICC by including cognitive and noncognitive elements. Third, it addresses the need to clarify relationships among dimensions. For example, despite strong validity evidence, the equally comprehensive cultural intelligence model (Earley & Ang, 2003) lacks theoretical explanations of the interplay between subdimensions. By basing the current model on a process model of individual behavior in complex social situations (Grossman et al., 2015), we highlight the dependent nature of the dimensions, implying a loose sequential relationship in which success in a later stage is dependent on the outcomes of an earlier stage. In sum, the present framework meets the three criteria (definition clarity, comprehensiveness, and subdimension relationship clarity) called for in the ICC literature.

Table 5 Task Types, Descriptions, and Potential Response Formats

Task Type	Description	Response formats
Cross-cultural scenario-based items	Participants will view videos or read about situations and respond to a series of questions. Questions may range from ranking the most useful information provided and evaluating appropriate behavioral responses (i.e., cultural knowledge application) to indicating likely perspectives of the individuals involved in the scenario (i.e., perspective taking).	Multiple-choice Short answer Likert-type Multiple selected-response
Comma switch	Individuals must retype a paragraph, swapping out the periods and the commas, after baseline typing performance has been assessed.	Text entry
Likely to be true	Based on a description of a fictitious character, individuals rate the likelihood of statements being true. Statements will range from directly related to the information (i.e., enjoying similar activities to ones suggested in the profile) to more stereotypical statements based on cultural membership.	Multiple-choice Short answer Likert-type
Spot the stereotype	Individuals read a paragraph and must select the sentences that are the most based on stereotypes.	Multiple-choice
Go/no-go	Individuals will respond to stimuli by clicking as directed in response to two stimuli.	Text entry
Flanker	Individuals will respond to stimuli by clicking as directed in response to stimuli.	Text entry
Emotional induction	Participants will be exposed to video clips to alter their mood; attitudes or skills could then be reassessed.	Likert-type Short answer
Troy et al. (2010) paradigm	Participants, prior exposure to a video clip designed to induce sadness, are instructed on an emotional regulation strategy. Emotion is measured before and after.	Likert-type Short answer
Incident recollection	Participants respond to prompts with a short written answer that is accessed using key word counts.	Short answer
Coaching task	Participants will be asked to resolve the cross-cultural difficulty or conflict experienced by a friend.	Selected-response Multiple selected-response (chat/nonchat based)
BASIC prompts	Individuals will respond to a variety of prompts, including statements (i.e., self-report items) and conditional reasoning questions.	Multiple-choice Forced-choice

Task Types and Response Formats

In crafting an ICC framework that entails assessing attitudes, cognitions, and behaviors, a complex assessment strategy will be necessary to adequately capture the content of each component. For that reason, a range of assessment considerations is presented in the following section, including task type and response option formats. Task type refers specifically to the type of activity, question, or prompt with which examinees would interact. Examples of these include SJTs or emotional induction. Response format refers to the format through which the response is communicated, such as short answer or multiple-choice. It should be noted that the tasks that we propose are not limited strictly to intercultural interactions, especially in the approach stage, as subdimensions such as tolerance of ambiguity are relevant in many situations in addition to intercultural interactions. However, when specifically measuring the ICC construct, tasks will explicitly reference elements of culture to best tap that domain. Table 5 contains an overview of the different task types and their potential response formats. Table 6 relates task type to the constructs of the present ICC model.

The next generation of ICC assessment requires more variety in task type. Historically, ICC has typically been assessed with self-report questions, in which the respondents report their own abilities, skill level, attitude, or knowledge. As discussed above, these commonly used self-report items may be appropriate for attitudinal constructs, but may be less so for

Table 6 Examples of Task Types to Assess Intercultural Competence

Item type	Approach				Analyze			Act	
	Positive cultural orientation	Tolerance of ambiguity	Cross-cultural self-efficacy	Self-awareness	Social monitoring	Perspective taking/suspending judgment	Cultural knowledge application	Behavior regulation	Emotion regulation
Self-report	X	X	X	X	X	X	X	X	X
Comma switch								X	X
Conditional reasoning	X	X	X						
SJT	X	X	X	X	X	X	X		
Likely to be true						X	X		
Spot the stereotypes						X	X		
Go/no-go								X	
Flanker task								X	
Troy et al. (2010) paradigm									X
Emotional induction									X
Cognitive process analysis	X	X	X	X	X	X	X	X	X

cognitive and behavioral skills. Considering the commonality of self-report items, assessment considerations are focused more heavily on these cognitive and behavioral dimensions. To that end, the following section discusses several task types and their associated response formats.

Intercultural Scenario-Based Items

Intercultural scenario-based (ICSB) items can be used to assess the appropriate behavioral response to a cross-cultural situation. ICSB items can be employed in the current context to focus on the specific skills of the framework, such as those in the analyze dimension. Potential questions in response to a situational passage or video could include those listed below. See Table 6 for a full list of the dimensions that could use the following item format:

1. What is the motivation of the first speaker? (perspective taking)
2. What additional information about the first speaker's culture would help you determine how to act? (cultural knowledge application)
3. Which of the following claims about the first speaker is likely to be true? (suspending stereotyped thinking)

Following the test or video that serves as the prompt for ICSB items, participants may be asked to respond using multiple-choice, Likert-type item, or short answer, each of which have strengths and weaknesses as response formats. Multiple-choice items allow multiple incorrect distractor options to be presented to the examinee, creating additional challenges in determining the correct answer. Likert-type items capture attitudinal constructs such as tolerance for ambiguity, as well as an individual's perceptions of their own abilities and their current emotional state in response to the situation. Short answer replies to open-ended questions allow for the most complex and qualitatively rich responses, in which participants generate their own unique responses. Finally, multiple items can address a single ICSB prompt, and different response formats could be used in conjunction with one another. It is important to note, however, that although the short answer response option might capture additional variance, items using this response option are resource intensive. They require the development of rubrics and two or more individuals to score written responses. However, advanced word recognition technology or other automated scoring procedures may remove the necessity of human scoring after the automated models have been validated. Although the technological development might require upfront resources, this could potentially decrease the cost of administering the assessment and the time required to score it.

One novel response format that might be used with ICSB task type involves the use of multiple selected responses. In other words, an examinee would be asked to select from two or more lists of options that explain their thinking or choices. For example, in response to a scenario, a participant could be asked to formulate an answer using three drop-down menus: one to indicate how he or she would feel in response to that scenario, a second to indicate what he or she would do, and a third to provide an explanation of choice. This method captures more information per scenario and allows participants to more precisely describe how they would respond to a situation. Moreover, it offers the potential to elicit more in-depth information from respondents without having to use constructed-response items that necessitate human scoring. The multiple drop-down menus can also be used in ICSB items to measure emotion regulation, a key component of the act stage. For example, in response to a scenario, participants can be asked how they would feel and what they would do to in response to those feelings. However, it should be noted that research on this response format may be less familiar to participants (Heerwegh & Loosveldt, 2002) and suffer from order effects (i.e., response options being selected based on place in the list; Couper, Tourangeau, Conrad, & Crawford, 2004).

Nontraditional Behavioral Skills Tests

Nontraditional behavioral skills tests (Gabrenya et al., 2012) represent another set of task types. Behavioral competencies such as flexibility, a key component of the act stage, may be captured by tasks such as those comprised by the Test of Attentional Performance battery (Zimmermann & Fimm, 2002). One of those tasks is the go/no-go task that requires participants to inhibit a response triggered by external stimuli. For example, an examinee may be asked to respond to go stimuli (e.g., a square in her screen) by pressing the space bar but refrain from pressing the key when she sees a circle (i.e., the no-go stimulus); the number of squares will far outweigh the number of circles, especially in the beginning, making pressing the space bar the dominant response. An individual's ability to withhold responding to the no-go stimulus,

assessed by the number of incorrect keystrokes (the number of space bar presses after seeing a circle), is used to assess behavioral inhibition (Simmonds, Pekar, & Mostofksy, 2008). Performance on this task may capture an important element of ICC: inhibiting the cultural response patterns from one's own culture and engaging in the norms of one's host culture. Go would be an appropriate option for the behavior regulation subdimension of the current model's act element. Additionally, several variants of this task exist (e.g., the Flanker task, which uses arrow keys; Koban & Pourtois, 2014). This range would allow for more variety in the task types presented to assessment takers. Participant reactions could also be captured as a way of assessing tolerance for ambiguity. Delays in response time after errors could also be captured as a way of measuring reaction to errors (Koban & Pourtois, 2014). In the context of ICC, higher sensitivity to error information could provide increased success. Concerns over lack of thematic continuity with the rest of the assessment could be addressed by embedding the basic task into a game set in against a fictitious cultural backdrop.

Nontraditional behavioral skills prompts would use text entry as a response format. This response format can capture behavioral responses; comparable to IATs that monitor speed and keyboard input, text entry could produce a skill-level score based on speed and incorrect keystroke. However, although this item format might be ideal for assessing the more difficult-to-capture skill dimensions (i.e., behavior regulation), it requires significant investment in development and pilot testing. Moreover, due to the novel nature of the examinee performance data generated by this response option, it is likely that normative performance data would be required to develop scoring guidelines. These items might also impose higher technological requirements on participants, both in terms of knowledge (i.e., computing ability) and equipment (i.e., more recent computers and faster internet connections). Finally, these approaches may be perceived to be unrelated to ICC by respondents due to salient differences in face validity.

Troy et al. Paradigm

Emotion regulation, the other subdimension of act, might also be measured in a nontraditional fashion using a recently developed paradigm (Troy, Wilhelm, Shallcross, & Mauss, 2010). The Troy et al. paradigm involves inducing a negative emotion in participants over a series of trials to assess emotion regulation skills. For the first induction, individuals view a video designed to trigger the desired emotion with no instructions; this trial serves as a baseline of emotional reactivity. Over subsequent inductions, individuals are given specific instructions to use a particular emotion regulation strategy (cognitive reframing: asking participants to think about the positive elements). The difference in reported emotion, as assessed by Likert-type items, is then used as a measure of emotion regulation ability. Results from Troy et al. (2010) suggest that it is a valid method (Gabrenya et al., 2012). Participants engaging in the emotion regulation strategy experienced less sadness than those who were given no instructions. To increase the thematic continuity of the assessment, the emotion-generating stimuli could be cross-cultural in nature (e.g., a filmed confrontation around cultural differences).

Response formats for the paradigm of Troy et al. (2010) include Likert-type and forced-choice items. Likert-type items offer the flexibility to assess a single emotion, but forced-choice items are by necessity comparative. In other words, forced-choice items would require creating potential response options that are of equal valence. If the aim of the task is only to assess sadness, than forced-choice items might be difficult to generate.

Conditional Reasoning

Conditional reasoning items represent another potential task type to assess ICC. Conditional reasoning items are designed to tap the unconscious and implicit elements of attitudes, and as such, are a good option when socially desirable responding is a concern. They examine cognitive biases under the pretense of an inductive reasoning exam. The respondent is presented with a scenario or choice of some sort and asked to pick from several response options that include a reason. Conditional reasoning items disguise the "right" answer—the options would include logic that appeals to the cognitive schema of individuals at all levels of the construct. For example, a conditional reasoning test item related to positive cultural attitude, an approach subdimension, could ask the examinee to select the reason for the increase in American car quality over the past 15 years after the introduction of foreign cars to American markets. Two of the options are as follows: "American companies have learned a lot from their international counterparts about quality manufacturing" and "American car manufacturers rose to the challenge in order to drive away foreign competition." To endorse the former option, an individual makes a cooperative assumption, but an individual endorsing the latter option expresses a more hostile and competitive option. A complete conditional reasoning test would score an individual's latent level of the construct based on the number

of times they endorsed the less positive options (C. M. Berry, Sackett, & Tobares, 2010). For measures of ICC attempting to assess general favorable attitudes toward culturally distinct others — essentially the inverse of ethnocentrism — the transparency of self-report items may preclude much variance. Beyond attitudes in the approach stage, these items might also be used to test the cognitive skills of the analyze stage as a standardized cognitive path analysis, in which individuals are asked to describe which way of knowing is closest to how they arrived at an answer. For example, response options would contain a clause that addresses the reasoning that supports the correct option. In other words, responses to an item could all describe the same behavioral response to the situation but have a different explanation for why that behavior was correct. Initial evidence suggests that these items reduce faking (LeBreton, Barksdale, Robin, & James, 2007); however, conditional reasoning items require extensive development efforts and pilot testing, making them a high-investment option.

The response format for conditional reasoning prompts could be a form of multiple choice that resembles the forced-choice response format. Each option presents an inference in reference to the prompt; two of the options contain framework-inconsistent inferences and serve only as distractors, one option reflects high levels of the target construct, and the fourth, low levels. The latter two response options are engineered to appeal or seem intuitive to an individual who has a high or low standing on that construct, respectively. An examinee must select one explanation to stand in for his or her reasoning in order to complete the task. Evidence supports this particular brand of multiple choice as being resistant to intentional faking (LeBreton et al., 2007).

Incident Recollection

Autobiographical incident recollection via advanced word recognition software or machine learning via keyword search can capture a variety of subdimensions. Individuals could be prompted to write short paragraphs about previous successful and unsuccessful cross-cultural experiences, or even theorize about what makes cross-cultural experiences successful, after which the automated scoring algorithm would look for keywords, phrases, and synonyms consistent with the proposed framework. Essay scoring options vary. For example, a score can be developed based on a frequency count of words related to specific skills (i.e., an analyze score created in part by the use of the words *viewpoint*, *perspective, what they were thinking, how they might consider it*, or *in their shoes*). An attitudinal score could be produced based on the overall valence (positivity – negativity) of the word choice. When paired with SJT stimuli, scoring the natural language of the respondent may be a productive method to assess whether their thought patterns map on to language consistent (or inconsistent, in the case of negative scoring) with the targeted constructs. This task type would rely primarily on the short answer response format, the benefits and drawbacks of which were previously discussed. Most notably, the short answer format is highly susceptible to faking, as participants could generate completely fictional accounts.

Coaching Task

For some testing situations, engendering specific emotions in the examinees may be considered inadvisable, especially negative emotions. In such cases, the following coaching paradigm might be used instead to test emotional regulation, the second subdimension of the act stage. Similar to ICSB items, these would describe a cultural situation in which a friend has experienced a negative situation, accompanied by a picture or short GIF when not video based. The correct answer would be a plausible response to the situation in combination with an emotion regulation strategy. Distractor options would include plausible responses that did not resolve the negative emotion expressed by the friend. Over several such items, it will be possible to assess an examinee's inclination toward emotion regulation. Although assessing this inclination is not the same as measuring an ability, it does provide the proxy measure intention, which has been shown to predict behavior (e.g., Ajzen, 1991).

This item type could, like conditional reasoning items, use the forced-choice response option. However, it could also use more novel and interactive response formats, in particular a chat-based selected-response format. This format would mimic a chat room environment but use a computer-directed avatar rather than a human-in-the-loop. Using computer-generated responses would reduce the cost while still creating an interactive examinee experience. However, developing items that use this format would require resource-intense investment initially. Such a format would facilitate a conversational tone. Participants could provide their advice and then be asked why they selected that advice option, providing an increased number of response combinations without necessitating an overwhelming number of response options within a single response list.

Additional Testing Considerations

Increased Psychological Fidelity

The assessment could also be adapted to replicate the cognitive and emotional complexity of real cross-cultural situations, a condition known as psychological fidelity. The inclusion of additional stimuli acknowledges the cognitive and emotional load present in cross-cultural interactions, which can be complex and challenging (Gabrenya et al., 2012). These stimuli could include foreign music (as a distraction), interrupting or competing tasks (increased cognitive load), or even minor emotional distress (e.g., a bad mood). This strategy would allow measurement conditions to more accurately reflect the conditions under which the skills assessed are used in reality and improve the assessment's ability to predict outcomes. They may also allow for the use of repeated measurement to tap other skills. For example, individuals could be asked to go through multiple rounds of the go/no-go task, with a negative mood induced in between rounds. Emotion regulation (part of act), could be assessed by the increase in errors in the second round.

Accessibility

In line with the best practices for testing established by the *Standards for Educational and Psychological Testing*, a next-generation assessment should be designed to "facilitate accessibility and minimize construct-irrelevant barriers for all test takers in the target population, as far as possible" (AERA et al., 2014, p. 57). The target population for this next-generation measure of ICC, American-based higher education students, is a diverse one; many universities have made great strides in accessibility for students with disabilities, funding for disadvantaged students, and attracting international students. Thus, a universal design (the principle of design in which products and environments are created to the maximal extent to be usable to everyone without needing case-by-case adaptation; Measured Progress & ETS Collaborative, 2012) should be considered. In short, as items are being crafted, test developers should aim to include aids and other considerations for examinees with differing abilities, language and cultural backgrounds, socioeconomic status, genders, and ages. For example, if the cultural scenarios are text-based prompts, reading level and working memory differences may impact examinees' scores. The use of visual aids such as charts and pictures may be incorporated to offset these demands and serve as memory cues, should video-based vignettes prove infeasible. These graphics could then also be accompanied by written descriptions for students with visual impairment. Additionally, efforts should be made to reduce the use of idiomatic language, which can serve as a barrier for examinees who speak English as a second language (Sireci, 2011). Further, some item types, such as the go/no-go task, require significant bandwidth and computational processing speed, and examinees' test-taking experience may then be adversely impacted by their lack of access to high-quality technology. The assessment could collect a baseline measurement by launching with a series of nonscored practice rounds so that technological differences might be taken into account for scoring purposes; a practice version would also serve as a tutorial to provide additional comfort to examinees with less exposure to such technology.

Conclusion

ICC has been identified as a critical life skill likely to predict success in the 21st century workforce. As universities begin to explore expanding traditional models of learning outcomes and emphasize these life skills, there is a need to assess whether students possess these critical competencies. In addition, assessments are needed to determine whether the abilities and skills underlying ICC improve during the university tenure of the student. Unfortunately, the current state of measurement of ICC leaves much to be desired, for several reasons. First, little consensus seems to exist regarding the requisite skills and abilities that contribute to ICC. Second, the measurement of ICC has overrelied on self-report methods that do not adequately cover the entire spectrum of the construct. Specifically, existing measures often tap self-referent cognitions without adequately capturing the affective and behavioral aspects that are inherent in intercultural interactions. Finally, the psychometric properties of existing measures leave much room for improvement. Although the reliabilities of existing measures meet professional standards, a relatively small number of studies provide evidence relating scores to other constructs, and even fewer provide evidence that the measures are related to outcomes of interest.

The three-pronged framework provided in this paper, approach, analyze, and act, is broad enough to cover important ICC construct domains, but also specific enough to result in clear operational definitions that can be used to guide the design of an ICC assessment. First, the framework assumes that ICC is an interactive process rather than treating the

construct as static. Second, the proposed framework follows this process through attitudinal, cognitive, and behavioral interactions that would likely occur in social cross-cultural communications. Finally, the framework is presented in a parsimonious fashion that enables clear interpretation of data that may result from a measure developed based on the framework. In addition to proposing a new framework, we deliberated on more innovative and interactive methods of assessing ICC that go beyond self-report. These methods have potential to improve the measurement of what has been an elusive construct, as well as to make the assessment experience enjoyable and insightful for students. It is our hope that the work presented in this paper will spur further discussion and examination of the ICC construct. In addition, we hope this continued discourse ultimately results in an operational measure of ICC that can assist higher education institutions in preparing a new generation of culturally competent global citizens.

References

Abbe, A., Gulick, L. M., & Herman, J. L. (2007). *Cross-cultural competence in Army leaders: A conceptual and empirical foundation*. Fort Leavenworth, KS: U.S. Army Research Institute for the Behavioral and Social Sciences, Leader Development Research Unit.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211.

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, *52*(1), 27–58.

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 Council on Education. (2016). *At home in the world toolkit*. Retrieved from https://www.acenet.edu/news-room/Pages/AHITW-Toolkit-Main.aspx

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

Anderson, P. H., & Lawton, L. (2011). Intercultural development: Study abroad vs. on-campus study. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *21*, 86–108.

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

Arasaratnam, L. A., & Doerfel, M. L. (2005). Intercultural communication competence: Identifying key components from multicultural perspectives. *International Journal of Intercultural Relations*, *29*(2), 137–163.

Arthur, W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored Internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment, 18*(1), 1–16.

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.

Association of American Colleges and Universities. (2011). *The LEAP vision for learning: Outcomes, practices, impact, and employers' views*. Washington, DC: Author.

Bazerman, M. H. (1990). *Judgment in managerial decision making*. New York, NY: Wiley.

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

Beechler, S., & Javidan, M. (2007). Leading with a global mindset. *Advances in International Management, 19*, 131–170.

Bennett, M. J. (1986). A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations, 10*(2), 179–196.

Bennett, M. J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.), *Education for the intercultural experience* (pp. 21–71). Yarmouth, ME: Intercultural Press.

Berry, C. M., Sackett, P. R., & Tobares, V. (2010). A meta-analysis of conditional reasoning tests of aggression. *Personnel Psychology, 63*, 361–384.

Berry, J. W., Kim, U., Power, S., Young, M., & Bujaki, M. (1989). Acculturation attitudes in plural societies. *Applied Psychology*, *38*, 185–206.

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.

Bikson, T. K., Treverton, G. F., Moini, J., & Lindstrom, G. (2003). *New challenges for international leadership: Lessons from organizations with global missions*. Santa Monica, CA: RAND. Retrieved from http://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1670.pdf

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

Bird, A., Mendenhall, M., Stevens, M. J., & Oddou, G. (2010). Defining the content domain of intercultural competence for global leaders. *Journal of Managerial Psychology, 25*(8), 810–828.

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

Black, J. S., Mobley, W. H., & Weldon, E. W. (2005). The mindset of global leaders: Inquisitiveness and duality. *Advances in Global Leadership, 4*, 181–200.

Black, J. S., & Stephens, G. K. (1989). The influence of the spouse on American expatriate adjustment and intent to stay in Pacific Rim overseas assignments. *Journal of Management, 15*, 529–544.

Boksem, M. S., Ruys, K. I., & Aarts, H. (2011). Facing disapproval: Performance monitoring in a social context. *Social Neuroscience, 6*, 360–368.

Brandl, J., & Neyer, A.-K. (2009). Applying cognitive adjustment theory to cross-cultural training for global virtual teams. *Human Resource Management, 48*(3), 341–353. doi:10.1002/hrm

Brenneman, M., Carney, L., Ezzo, C., Klafehn, J., Kyllonen, P., Burrus, J., … , Gallus, J. A. (2016). *Development of an assessment of cross-cultural competence: An expanded literature review* (contract number W5J9CQ-12-C-0039, Army Research Institute Technical Report). Manuscript in preparation.

Brown, A., & Maydeu-Olivares, A. (2011). *Forced-choice Five Factor markers*. Retrieved from PsycTESTS. doi:10.1037/t05430-000

Bücker, J., & Poutsma, E. (2010). Global management competencies: A theoretical foundation. *Journal of Managerial Psychology, 25*, 829–844.

Bureau of Educational and Cultural Affairs. (2013). *Fulbright fact sheet.* Retrieved from http://eca.state.gov/files/bureau/fulbright_fact_sheet_2.pdf

Byram, M. (1997), *Teaching and assessing intercultural communicative competence*. Clevedon, England: Multilingual Matters.

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

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.

Chaney, S. K., & Christiansen, N. D. (2004, April). *Disentangling applicant faking from personality: Using covariance to detect response distortion*. Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

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: A synthesis. *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.

Christiansen, N. D., Burns, G. N., & Montgomery, G. E. (2005). Reconsidering forced-choice item formats for applicant personality assessment. *Human Performance, 18*(3), 267–307.

Clemson University. (2016). *General education competencies.* Retrieved from http://www.clemson.edu/academics/programs/eportfolio/general-education.html

Conway, J. L. (2008). *An analysis of intercultural competence training for community college employees* (Doctoral dissertation). Retrieved from ProQuest Digital Dissertations (AAT 3304562).

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

Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004). What they see is what we get: Response options for web surveys. *Social Science Computer Review, 22*(1), 111–127.

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

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

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.

Deardorff, D. K. (2004). Internationalization: In search of intercultural competence. *International Educator, 13*(3), 13–15.

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

Deardorff, D. K. (Ed.). (2009). *The SAGE handbook of intercultural competence*. Thousand Oaks, CA: SAGE.

De Haan, H. (2014). Can internationalisation really lead to institutional competitive advantage? – A study of 16 Dutch public higher education institutions. *European Journal of Higher Education, 4*(2), 135–152. doi:10.1080/21568235.2013.860359

DeJaeghere, J. G., & Cao, Y. (2009). Developing U.S. teachers' intercultural competence: Does professional development matter? *International Journal of Intercultural Relations, 33*(5), 437–447.

DeJaeghere, J. G., & Zhang, Y. (2008). Development of intercultural competence among US American teachers: Professional development factors that enhance competence. *Intercultural Education, 19*(3), 255–268. doi:10.1080/14675980802078624

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

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*, 1–23.

Earley, P. C., & Ang, S. (2003). *Cultural intelligence: Individual interactions across cultures*. Stanford, CA: Stanford University.

Earley, P. C., & Peterson, R. S. (2004). The elusive cultural chameleon: Cultural intelligence as a new approach to intercultural training for the global manager. *Academy of Management Learning & Education*, *3*(1), 100–115.

Eisenberg, J., Lee, H. J., Brück, F., Brenner, B., Claes, M. T., Mironski, J., & Bell, R. (2013). Can business schools make students culturally competent? Effects of cross-cultural management courses on cultural intelligence. *Academy of Management Learning & Education, 12*, 603–621.

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.

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. (1995). Language, culture, and world view: Exploring the nexus. *International Journal of Intercultural Relations, 19*, 143–153.

Fantini, A. E. (2009). Assessing intercultural competence: Issues and tools. In D. K. Deardorff, (Ed.), *The SAGE handbook of intercultural competence* (pp. 456–476). Thousand Oaks, CA: SAGE.

Fantini, A. E., Arias-Galicia, F., & Guay, D. (2001). *Globalization and 21st century competencies: Challenges for North American higher education*. Boulder, CO: Western Interstate Commission for Higher Education.

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

Fellows, K. L., Goedde, S. D., & Schwichtenberg, E. J. (2014). What's your CQ? A thought leadership exploration of cultural intelligence in contemporary institutions of higher learning. *Romanian Journal of Communication and Public Relations/Revista Română de Comunicare şi Relaţii Publice, 2*, 13–34.

Fennes, H., & Hapgood, K. (1997). *Intercultural learning in the classroom: Crossing borders*. London, England: Cassell.

Fischer, R. (2011). Cross-cultural training effects on cultural essentialism beliefs and cultural intelligence. *International Journal of Intercultural Relations, 35*(6), 767–775. doi:10.1016/j.ijintrel.2011.08.005

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

Gabrenya, W. K., Jr., Griffith, R. L., Moukarzel, R. G., Pomerance, M. H., & Reid, P. (2012). Theoretical and practical advances in the assessment of cross-cultural competence. In D. D. Schmorrow & D. M. Nicholson (Eds.), *Advances in design for cross-cultural activities: Part I* (pp. 317–327). Boca Raton, FL: CRC Press.

Gallois, C., Franklyn-Stokes, A., Giles, H., & Coupland, N. (1988). Communication accommodation in intercultural encounters. In Y. Y. Kim & W. B. Gudykunst (Eds.), *Theories in intercultural communication* (pp. 157–158). Newbury Park, CA: Sage.

Global Leadership Excellence. (2010). Retrieved May 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.

Graf, A., & Mertesacker, M. (2009). Intercultural training: Six measures 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

Griffith, D. A., & Harvey, M. G. (2000). An intercultural communication model for use in global interorganizational networks. *Journal of International Marketing, 9*(3), 87–103.

Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*(3), 341–355.

Griffith, R. L., & Converse, P. D. (2011). The rules of evidence and the prevalence of applicant faking. *New Perspectives on Faking in Personality Assessment*, *1*, 34–52.

Griffith, R. L., Malm, T., English, A., Yoshita, Y., & Gujar, A. (2006). Applicant faking behavior: Teasing apart the influence of situational variance, cognitive biases, and individual differences. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 151–177). Greenwich, CT: IAP.

Griffith, R. L., & Peterson, M. H. (2008). The failure of social desirability measures to capture applicant faking behavior. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*(4), 308–311.

Gross, J., Salovey, P., Rosenberg, E. L., & Fredrickson, B. L. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271–299.

Grossman, R., Thayer, A. L., Shuffler, M. L., Burke, C. S., & Salas, E. (2015). Critical social thinking: A conceptual model and insights for training. *Organizational Psychology Review, 5*(2), 99–125.

Gudykunst, W. B. (Ed.). (2003). *Cross-cultural and intercultural communication*. Thousand Oaks, CA: SAGE.

Gudykunst, W. B., Matsumoto, Y., Ting-Toomey, S., Nishida, T., Kim, K., & Heyman, S. (1994). *Measuring self construals across cultures*. Paper presented at the annual meeting of the International Communication Association, Sydney, Australia.

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). A test of anxiety/uncertainty management theory: The intercultural adaptation context. *Communication Quarterly, 46*, 309–326.

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

Hao, R. N. (2012). A critical review of intercultural communication research. In N. Bardhan & M. P. Orbe (Eds.), *Identity research and communication: Intercultural reflections and future directions* (pp. 71–86). Lanham, MD: Lexington Books.

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.

Hart Research Associates. (2015). *Falling short? College learning and career success.* Washington, DC: Author. Retrieved from Association of American Colleges & Universities website http://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf

Haslberger, A., Brewster, C., & Hippler, T. (2013). The dimensions of expatriate adjustment. *Human Resource Management, 52*(3), 333–351.

Hayes, D. J., Shealy, C. N., Sivo, S. A., & Weinstein, Z. C. (1999, August). *Psychology, religion, and Scale 5 (Religious Traditionalism) of the "BEVI."* Poster session presented at the meeting of the American Psychological Association, Boston, MA.

Heerwegh, D., & Loosveldt, G. (2002). An evaluation of the effect of response formats on data quality in web surveys. *Social Science Computer Review*, *20*(4), 471–484.

Hofstede, G. 1980. *Culture's consequences: International differences in work-related values.* Beverly Hills, CA: SAGE.

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

Howard-Hamilton, M. F., Richardson, B. J., & Shuford, B. (1998). Promoting multicultural education: A holistic approach. *College Student Affairs Journal*, *18*(1), 5–17.

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

Imahori, T. T., & Lanigan, M. L. (1989). Relational model of intercultural communication competence. *International Journal of Intercultural Relations, 13*, 269–286.

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

Institute of International Education. (2015). *Open Doors report on international educational exchange.* Retrieved from http://www.iie.org/Research-and-Publications/Open-Doors#.VxEsVDArKUk

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.

Jackson, D. N., Wroblewski, V. R., & Ashton, M. C. (2000). The impact of faking on employment tests: Does forced choice offer a solution? *Human Performance, 13*(4), 371–388.

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

Jarrell, K., Alpers, R. R., Brown, G., & Wotring, R. (2008). Using BaFa' BaFa' in evaluating cultural competence of nursing students. *Teaching and Learning in Nursing, 3*(4), 141–142.

Jauregui, M. (2013). *Cross-cultural training of expatriate faculty teaching in international branch campuses* (Unpublished doctoral dissertation). University of Southern California, Los Angeles.

Johnson, T. P., Shavitt, S., & Holbrook, A. L. (2011). Survey response styles across cultures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 130–175). Cambridge, England: Cambridge University Press.

Kahr-Gottlieb, D., & Papst, P. (2013). Competence training for medical students in Austria. *European Journal of Public Health, 23*. Retrieved from http://eurpub.oxfordjournals.org/content/eurpub/23/suppl_1/local/complete-issue.pdf

Kaufmann, H. R., Englezou, M., & García-Gallego, A. (2014). Tailoring cross-cultural competence training. *Thunderbird International Business Review, 56*(1), 27–42.

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

Kim, Y. Y. (2000). *Becoming intercultural: An integrative theory of communication and cross-cultural adaptation*. Thousand Oaks, CA: SAGE.

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–207.

King, P. M., & Baxter Magolda, M. B. (2005). A developmental model of intercultural maturity. *Journal of College Student Development, 46*(6), 571–592.

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

Kline, P. (Ed.). (2000). *A psychometrics primer*. London, England: Free Association Books.

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

Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology, 93*(1), 140–154.

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.

Lambert, R. D. (Ed.). (1994). *Educational exchange and global competence*. New York, NY: Council on International Educational Exchange.

Lane, H. W., Maznevski, M. L., & Mendenhall, M. (2004). Globalization: Hercules meets Buddha. In H. W. Lane, M. Maznevski, M. E. Mendenhall, & J. McNett (Eds.), *The Blackwell handbook of global management: A guide to managing complexity* (pp. 3–25). Malden, MA: Blackwell.

LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology, 92*(1), 1–16.

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

Levy, O., Beechler, S., Taylor, S., & Boyacigiller, N. A. (2007). What we talk about when we talk about 'global mindset': Managerial cognition in multinational corporations. *Journal of International Business Studies, 38*, 231–258.

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 of World Business*, *50*(1), 3–14.

Lodder, G. A., Scholte, R. J., Goossens, L., Engels, R. E., & Verhagen, M. (2016). Loneliness and the social monitoring system: Emotion recognition and eye gaze in a real-life conversation. *British Journal of Psychology*, *107*(1), 135–153. doi:10.1111/bjop.12131

Lustig, M. W., & Koester, J. (2003). *Intercultural competence: Interpersonal communication across cultures*. Boston, MA: Allyn and Bacon.

Martin, B. A., Bowen, C.-C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32*, 247–256.

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.

McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730–740.

McFarland, L. A., & Ryan, A. M. (2006). Toward an integrated model of applicant faking behavior. *Journal of Applied Social Psychology, 36*(4), 979–1016.

Measured Progress & ETS Collaborative. (2012). *Smarter Balanced Assessment Consortium: English language arts item and task specifications*. Retrieved from http://www.cccoe.k12.ca.us/edsvcs/commoncore/ELAGeneralItemandTaskSpecifications.pdf

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

Messner, W., & Schäfer, 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.

NAFSA: Association of International Educators. (2016). *NAFSA International Student Economic Value Tool.* Retrieved from http://www.nafsa.org/Explore_International_Education/Impact/Data_And_Statistics/NAFSA_International_Student_Economic_Value_Tool

Navas, M., García, M. C., Sánchez, J., Rojas, A. J., Pumares, P., & Fernández, J. S. (2005). Relative Acculturation Extended Model (RAEM): New contributions with regard to the study of acculturation. *International Journal of Intercultural Relations*, *29*(1), 21–37.

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

Paige, R. M. (1993). *Education for the intercultural experience.* New York, NY: Nicholas Brealey.

Paige, R. M., & Goode, M. L. (2009). Intercultural competence in international education administration. Cultural mentoring: International education professionals and the development of intercultural competence. In D. K. Deardorff (Ed.), *The SAGE handbook of intercultural competence* (pp. 333–349). Thousand Oaks, CA: SAGE.

Parkins, G. A., & Williams, E. (2011). The big picture: SWTG introduces the comprehensive training environment. *Special Warfare: The Professional Bulletin of Army Special Operations, 24*(2), 15-17.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598–609.

Peterson, M. H., & Griffith, R. L. (2006). Faking and job performance: A multifaceted issue. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 233–262). Greenwich, CT: IAP.

Peterson, M. H., Griffith, R. L., Isaacson, J. A., O'Connell, M. S., & Mangos, P. M. (2011). Applicant faking, social desirability, and the prediction of counterproductive work behaviors. *Human Performance, 24*, 270–290.

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), 615–621.

Pusch, M. (1994). The chameleon capacity. In R. D. Lambert (Ed.), *Educational exchange and global competence* (pp. 205–210). New York, NY: Council on International Educational Exchange.

Ramsey, J. R., Barakat, L. L., & Aad, A. A. (2014). Commitment to the study of international business and cultural intelligence: A multilevel model. *Journal of Teaching in International Business, 25*, 267–282.

Rathje, S. (2007). Intercultural competence: The status and future of a controversial concept. *Language and Intercultural Communication*, *7*(4), 254–266.

Reid, P., Kaloydis, F., Sudduth, M., & Greene-Sands, A. (2012). *Executive Summary: A framework for understanding cross-cultural competence in the Department of Defense* (DEOMI Technical Report 15–12). Patrick AFB, FL: Defense Equal Opportunity Management Institute.

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.

Rosenblatt, V., Worthley, R., & MacNab, B. (2013). From contact to development in experiential cultural intelligence education: The mediating influence of expectancy disconfirmation. *Academy of Management Learning & Education*, *12*(3), 356–379.

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

Sackett, P. R. (2011). Integrating and prioritizing theoretical perspectives on applicant faking of personality measures. *Human Performance 24*(4), 379–385.

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-%20Understanding%20the%20Cultural%20Orientations%20Approach.12.06.2012.pdf

Scott, P. (2000). Globalisation and higher education: Challenges for the 21st century. *Journal of Studies in International Education, 4*(1), 3–10. doi:10.1177/102831530000400102

Shaffer, M. S. (Ed.). (2012). *Public culture: Diversity, democracy, and community in the United States*. Philadelphia: University of Pennsylvania.

Shealy, C. N. (2004). A model and method for "making" a combined-integrated psychologist: Equilintegration (EI) theory and the Beliefs, Events, and Values Inventory (BEVI). *Journal of Clinical Psychology*, *60*(10), 1065–1090.

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 in multicultural teams. In S. Ang & L. Van Dyne (Eds.), *Handbook of cultural intelligence: Theory, measurement, and applications* (pp. 177–191). Abington, England: M. E. Sharpe.

Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of go/no-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia*, *46*(1), 224–232.

Sireci, S. G. (2011). Evaluating test and survey items for bias across languages and cultures. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 216–240). Cambridge, MA: Cambridge University Press.

Skinner, M. (2002). The renaissance of unconventional warfare as an SF mission. *Special Warfare, 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.

Soubelet, A., & Salthouse, T. A. (2011). Influence of social desirability on age differences in self-reports of mood and personality. *Journal of Personality, 79*, 741–762.

Spitzberg, B. H., & Changnon, G. (2009). Conceptualizing intercultural competence. In D. K. Deardorff (Ed.), *The SAGE handbook of intercultural competence* (pp. 2–52). Thousand Oaks, CA: SAGE.

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.

Taylor, S. E. (1989). Escape from reality: Illusions in everyday life. In B. M. Staw, (Ed.), *Psychological dimensions of organizational behavior* (pp. 131–155). Upper Saddle River, NJ: Pearson/Prentice Hall.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210.

Teräs, M., & Lasonen, J. (2013). The development of teachers' intercultural competence using a change laboratory method. *Vocations and Learning, 6*(1), 107–134.

Terrell, S. R., & Rosenbusch, K. (2013). How global leaders develop. *Journal of Management Development, 32*(10), 1056–1079.

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.

Ting-Toomey, S. (1999). *Communicating across cultures*. New York, NY: Guilford Press.

Ting-Toomey, S., & Kurogi, A. (1998). Facework competence in intercultural conflict: An updated face-negotiation theory. *International Journal of Intercultural Relations, 22*, 187–225.

Troy, A. S., Wilhelm, F. H., Shallcross, A. J., & Mauss, I. B. (2010). Seeing the silver lining: Cognitive reappraisal ability moderates the relationship between stress and depressive symptoms. *Emotion*, *10*(6), 783–795.

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 global manager: An empirical study on the role of academic international sojourns. *Academy of Management Learning & Education, 13*, 187–207.

Visweswaran, C., & Ones, D. S. (1999). Meta-analysis of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210.

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.

Williams, T. R. (2005). Exploring the impact of study abroad on students' intercultural communication skills: Adaptability and sensitivity. *Journal of Studies in International Education*, *9*(4), 356–371.

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

Yamaguchi, Y., & Wiseman, R. L. (2001). *Locus of control, self construals, intercultural effectiveness, and cross-cultural adjustment*. Paper presented at the annual meeting of the International Communication Association, Washington, DC.

Zhang, Y. (2012). Expatriate development for cross-cultural adjustment: Effects of cultural distance and cultural intelligence. *Human Resource Development Review*, *12*(2), 177–199. doi:10.1177/1534484312461637

Ziegler, M., MacCann, C., & Roberts, R. D. (2011). Faking: Knowns, unknowns, and points of contention. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 3–16). Oxford, England: Oxford University Press.

Zimmermann, P., & Fimm, B. (2002). A test battery for attentional performance. In A. H. van Zomeren, M. Leclercq, & P. Zimmermann (Eds.), *Applied neuropsychology of attention: Theory, diagnosis and rehabilitation* (pp. 110–151). New York, NY: Psychology Press.

Appendix

Literature Search Strategy

In order to conduct a comprehensive review of the literature, an iterative search process was implemented. As of now, the EBSCO database and *Google Scholar* were the primary databases used to obtain relevant articles. Keywords used in this search included *intercultural competence, cross-cultural competence, college students, university students, postsecondary education, higher education.*

Suggested citation:

Griffith, R. L., Wolfeld, L., Armon, B. K., Rios, J., & Liu, O. L. (2016). *Assessing intercultural competence in higher education: Existing research and future directions* (Research Report No. RR-16-25). Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12112