Journal of Cross-Cultural Psychology
44(6) 849–873

DOI: 10.1177/0022022113492891
jccp.sagepub.com
SAGE

Assessing Cross-Cultural Competence: A Review of Available Tests

David Matsumoto[1,2] and Hyisung C. Hwang[1,2]

Abstract

Recent years have witnessed the emergence of a number of tests that measure cross-cultural competence; yet to date there is no review of their validity and reliability. This article addresses this gap in the literature. We discuss issues associated with evaluation of the content, construct, and ecological validity of such tests, and review the evidence for 10 tests. We evaluate that evidence, draw conclusions about the tests with the best evidence for ecological validity, and provide recommendations for future research in this area.

Keywords

measurement/statistics, methodology, intelligence/abilities

One important area of research related to cross-cultural competence (3C) involves the development and validation of tests to assess it, which has theoretical import and practical ramifications. Theoretically, they can help to identify the psychological constructs necessary for intercultural adaptation and adjustment, aiding in the creation of models that improve our understanding of this rich and complex phenomenon. Practically such tests can identify goals of intervention, allowing practitioners to design effective training programs and assess efficacy, which are important for organizations and individuals.

The purpose of this article is to provide a comprehensive review of the available tests of 3C. Although initial efforts to create such tests started decades ago, recently multiple tests have emerged in the literature. To our knowledge, there has not been a review of them published in a peer-reviewed journal; this article fills that gap.

We begin by describing issues associated with the creation of tests of 3C, which provide the basis for evaluating their content validity, and then issues associated with establishing their psychometric reliability and validity, which provide the basis for an evaluation of their construct and ecological validities. We review the psychometric evidence of ten 3C tests, aggregating information on each heretofore not assembled in a single place in the literature. We evaluate that evidence, draw conclusions, and give recommendations about future research based on that evaluation.

[1]San Francisco State University, CA, USA
[2]Humintell LLC, Berkeley, CA, USA

Corresponding Author:
David Matsumoto, Department of Psychology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA.
Email: dm@sfsu.edu

Methods for Creating Tests of 3C and Content Validity

Creating a test of 3C typically begins with the identification of the desirable outcomes to be predicted, the target cultures within which competence is to be demonstrated, and the knowledge, skills, abilities, and other (KSAOs) factors that are necessary to demonstrate competence. Based on this analysis, initial item pools that assess the hypothesized KSAOs are created. The quality of this process is the basis by which evaluations of a test's content validity can occur. Below, we discuss briefly each of these issues as they relate to this review.

Identifying Desirable Outcomes

In this literature, outcomes are broadly referred to as adaptation and adjustment. These terms can have different meanings to different researchers and are sometimes used interchangeably. Thus, we make explicit here our definitions of them.

Adaptation is the process of altering one's behavior in response to the environment, circumstances, or social pressure. Changing which side of the street on which to drive when going from England to France, for instance, is an alteration in behavior in response to different environments, just as learning to use chopsticks when in East Asia. In the literature, adaptation has been assessed by management styles, leadership behaviors, performance in culturally diverse teams, vocational interests, international orientation, relationship quality, interactive behaviors, and so forth.

Adjustment refers to the subjective experiences associated with adaptation, and may be assessed by mood states, self-esteem, self-awareness, physical health, self-confidence, stress, psychological and psychosomatic concerns, early return to one's home country, dysfunctional communication, culture shock, depression, anxiety, diminished school and work performance, and difficulties in interpersonal relationships. In extreme cases, negative adjustment can involve antisocial behavior (gangs, substance abuse, crime) and even suicide.

Successful intercultural adaptation and adjustment involves the adoption of behaviors that accomplishes goals and achieves tasks while at the same time minimizing negative adjustment outcomes and maximizing positive ones. This includes having successful relationships with people from other cultures; feeling that interactions are warm, cordial, respectful, and cooperative; accomplishing tasks in an effective and efficient manner; and managing psychological stress effectively in one's daily activities, interpersonal relations, and work environment (Black & Stephens, 1989; Brislin, 1993; Gudykunst, Hammer, & Wiseman, 1977).

Identifying Cultures

Developers of 3C tests need to decide which cultures are the targets within which to demonstrate competence and obtain desirable outcomes. Broadly speaking, there are two approaches. A culture-specific approach identifies specific cultures or regions in which individuals must demonstrate competence, and a culture-specific 3C test likely contains culture-specific item content. A culture-general approach is based on the assumption that individuals inherently possess KSAOs related to 3C without regard to a specific culture or region. These characteristics constitute an internal psychological resource pool that individuals tap wherever they are. There are hybrids of these approaches: a culture-specific test, for example, may assess constructs that can evolve into a culture-general test and vice versa. This article focuses on culture-general tests.

Identifying KSAOs

Once desirable outcomes and target cultures are identified, researchers need to identify the KSAOs required to produce the desired outcomes in those cultures. These can come

from previous theory, research, or experience. Their breadth can vary greatly. On one hand, if a developer is interested in testing whether specific constructs (e.g., openness, flexibility) predict desirable outcomes, there may be a focus on just those constructs and others directly related to them. On the other hand, if a developer does not care which constructs predict the desirable outcomes, then there would be a much broader assessment of the possible KSAOs.

Generating Item Pools

Once KSAOs are identified, the next step is to generate initial item pools that operationalize them. The most common approach to test development is to create questionnaires with scalar response items (one exception is the Intercultural Behavioral Assessment [IBA]/Behavioral Assessment Scale for Intercultural Communication Effectiveness [BASIC], described below). The items may either be modifications of items from existing tests that assess similar constructs or created anew. Initial versions of a test typically include large item pools and items are eliminated from the initial pool in the validation process, during which researchers balance desires for higher reliability of measurement with practicality, resulting in final item pools that allow for reasonably reliable measurement of KSAOs while not being too long.

Criteria for Evaluating Content Validity

The quality of the process described above forms the basis of evaluations of content validity, which address the following questions:

1. Were the desirable outcomes clearly identified and defined?
2. Were the target cultures clearly identified?
3. Were the KSAOs associated with the desirable outcomes in the target cultures clearly defined?
4. Did the KSAOs exhaust the possibilities of all KSAOs that could potentially be predictive of the desirable outcomes?
5. Did the generated item pools exhaust the possible universe of measurement for each KSAO?

Methods for Establishing Psychometric Reliability and Validity of Tests of 3C: Construct and Ecological Validity

Definition of Terms

Because there are differences in usage of terms associated with reliability and validity, we make explicit here our use of them. We define *construct validity* as the verification that the test measures the constructs it was designed to measure. There are several ways construct validity can be established. Confirming the underlying latent structure of the items, either through Exploratory Factor Analysis (EFA), Principal Components Analysis (PCA), or Confirmatory Factor Analysis (CFA), demonstrates structural validity, a type of construct validity. Establishing relationships with other psychological constructs associated to those assessed in the 3C test demonstrates convergent validity, another type of construct validity. Demonstrating intercorrelations among scales of a multiple-scale 3C test is another form of convergent validity. Demonstrating that the 3C test is associated with other psychological constructs that other 3C tests are not is divergent validity, which is a type of construct validity.

Ecological validity refers to the documentation that the 3C test predicts measures of desired outcomes that serve as criterion variables, that is, measures of intercultural adjustment,

adaptation, communication competence, interaction success, and so forth. Ecological validity can be established in several ways. One is to demonstrate associations between the proposed 3C test and measures of the criterion variables. If the criterion variables are obtained at the same time as the 3C test, we consider that concurrent ecological validity; if obtained later, we consider that predictive ecological validity. Ecological validity can also be assessed by demonstrating changes in pre–post scores in studies examining the efficacy of intercultural training or the effects of sojourns, or by extreme group difference tests, such as between known groups of individuals who are interculturally competent and those who are not. Demonstration that the 3C test predicts a criterion variable above and beyond other tests provides evidence for incremental ecological validity. In the literature, researchers have labeled the various types of ecological validity described here as predictive validity, external validity, or criterion validity; here we label them as ecological validity.

Two Approaches to Validation

Once an initial item pool is generated, broadly speaking, there are two approaches available to reduce and refine it (Anastasi & Urbina, 1997), which have been utilized in this area of study. They differ not in what is done, but instead in the order in which things are done at the initial stages of validation. The *construct validity–driven approach* involves first the identification of the latent structure underlying the item pool, typically through EFA or PCA, and item reduction by removing items not associated to the latent structure. Ecological and further construct validity tests are performed on the resulting latent structures after initial item reduction has occurred. The *ecological validity–driven approach* involves first testing the ecological validity of individual items of the initial item pool and item reduction by removing items not associated with criterion variables. A latent structure can subsequently be generated, and further tests of construct and ecological validity can be established using either the surviving set of items or the latent structure.

The construct validity–driven approach is the more common procedure for psychological tests. Its advantage is that it is likely to generate clearer mental constructs assessed by the latent structures and measurement models, exemplified by larger percentages of cumulative variance accounted for in EFAs, more clearly interpretable factor structures, and more internally reliable scale scores. This method is also likely to produce clearer pictures of the nomological network with other constructs. A potential disadvantage, however, is that the resulting factors may not be as robust as possible in predicting criterion variables across a wide range of samples or methodologies because the initial focus is on the latent structure and not on ecological validity. The advantage of the ecological validity–driven approach is that it is more likely to generate items that are more robust in predicting criterion variables; its disadvantage is that the surviving items are less likely to generate clear latent structures. Thus, scale scores are more likely to have lower internal reliabilities.

Criteria for Evaluation

Regardless of the specific approach taken, we believe that the main criterion against which tests of 3C should be evaluated is the strength of the evidence for ecological validity, which requires addressing the following issues:

1. *Validity and reliability of the criterion variables*: Criterion variables need to be reasonably desirable outcomes in relation to 3C, commensurate with researchers' theoretical frameworks, and, most importantly, validly and reliably operationalized.
2. *The number and breadth of cross-cultural samples*: Tests of 3C need to be validated in different cultures. The greater the number of cross-cultural samples, the better; the greater use of nonstudent samples, the better.

3. *Mixed methodologies*: Examining correlations between a questionnaire-based 3C test and other questionnaires is a common practice, but is limited by concerns about halo and common method variance. Ecological validity tests that involve behavioral tasks, interviews, participation in intercultural training or sojourns, and other non-questionnaire-based assessments bolster the case for ecological validity.
4. *Time of assessment*: Although concurrent validity tests are important, predictive validity tests are also important, especially given the importance and practical utility of a 3C test to be used as a predictor of future intercultural success.
5. *Incremental validity*: 3C tests should demonstrate that they predict outcomes above and beyond what is already predicted by other 3C and non-3C tests.

Method

We searched for relevant tests through Psycinfo, Google Scholar, and search engines of primary journals publishing in the cross-cultural and intercultural areas in psychology and business/management. Sources were obtained and tests mentioned in the sources that had not originally been uncovered in the searches were researched manually. In several cases (BASIC, Cultural Intelligence Scale [CQ], IBA, Intercultural Sensitivity Inventory [ICSI], Intercultural Development Inventory [IDI], and Multicultural Personality Inventory [MPQ]), originating authors of the tests were contacted and requested to provide references concerning the documentation of the validity and reliability of the test. Tests were retained for review if they met the following criteria:

1. The test attempted to predict outcomes related to successful adjustment or adaptation to new cultural environments such as international sojourns or deployments, or living or working in multicultural environments with people from cultures different than oneself. We did not, therefore, include the many tests and measures developed in the domain of cross-cultural counseling or therapy (e.g., the Cross-Cultural Counseling Inventory, LaFromboise, Coleman, & Gerton, 1993; Scale of Ethnocultural Empathy, Wang et al., 2003), because a proper review of those tests and measures should occur within a review of the KSAOs associated with therapists and counselors (see reviews by Gamst, Liang, & Der-Karabetian, 2011; Ponterotto, Rieger, Barrett, & Sparks, 1994). We also did not consider the many tests and measures associated with the assessment of culture-related attitudes, values, beliefs, or abilities not directly assessing 3C KSAOs (e.g., the Munroe Multicultural Attitude Scale Questionnaire, Munroe & Pearson, 2006; the Schwartz Values Scale, Schwartz, 1992; the Personal Orientation Inventory, Uhes & Shybut, 1971).
2. The test was designed for multiple uses and with efforts at demonstrating its psychometric properties, and was not a measure designed for a single study (e.g., the Gesture Recognition test in Molinsky, Krabbenhoft, Ambady, & Choi, 2005; the Test of Intercultural Sensitivity used in Weldon, Carlston, Rissman, Slobodin, & Triandis, 1975).
3. The test was based on a culture-general approach rather than a culture-specific approach.
4. Sources documenting the psychometric properties of the test were published in empirical articles in peer-reviewed journals in English. We relied on empirical articles that reported primary data as opposed to reviews of data or theoretical models, as empirical articles should be the primary data sources of a formal review of the psychometric properties of tests. We relied on peer-reviewed articles because they provided a standard of quality control over the information presented. We did not include reviews of a test even though they were published in a peer-reviewed journal and even though they reported data not reported elsewhere because these generally did not provide the methodological detail typical of original data reports and thus could not be evaluated for quality control (e.g., van Oudenhoven, Timmerman, & van der Zee, 2007). We also did not include information

obtained in books, chapters, user manuals, technical documents, unpublished manuscripts, or reports produced by government or private industry for the same reason. This criterion also resulted in the nonconsideration of tests of 3C developed as business tools because most are not associated with publications in peer-reviewed journals that document their psychometric properties (see Fantini, 2009, for a review of many of these).

Application of the criteria above resulted in the selection of 10 tests for review. Below, we first describe the procedures used for test creation and then the evidence for construct and then ecological validity. Later, we summarize and then evaluate that evidence.

Review of the Available 3C Tests

Cross-Cultural Adaptability Inventory (CCAI)

Kelley and Meyers (1987) developed the CCAI based on the available literature and consultation with subject matter experts, creating a preliminary list of 59 items that tapped traits and skills associated with the ability to adapt effectively to other cultures. Twenty-five cross-cultural trainers and consultants rated the importance of each item, and their ratings were compared with the cross-cultural literature; 16 items were consistently rated the highest. Items were grouped into four skill sets based on the correlations among the items and relevant discussions in the literature. Four judges also sorted the items into four categories. A fifth skill set (positive regard for others) was added to the list. Ten items for each of the five skill sets were written and focus groups were used to gather feedback about the items. Then 653 cross-cultural trainers rated the pool, and the fifth dimension was removed and items were moved from one dimension to another. The final CCAI assesses four dimensions: Emotional Resilience (18 items), Flexibility/Openness (15 items), Perceptual Acuity (10 items), and Personal Autonomy (7 items).

There is mixed evidence for the construct validity of the CCAI. Montagliani and Giacalone (1998) administered it to 35 employees from U.S.-based international corporations and 77 undergraduates enrolled in an international management course, along with Self-Monitoring Scale (Snyder, 1974) and the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 2002), which produces three scores: Impression Management, Self-Deceptive Enhancement, and a total score. Scores on the BIDR were modestly correlated with the CCAI total score (*r*s = .23, .31, and .29). Davis and Finney (2006) administered the CCAI to 709 U.S. university undergraduates, but CFAs indicated that the fit of the four-factor model was "very poor" (p. 323) and there were very high correlations among the four factors (.87-.98), suggesting problems in their discriminant validity; EFA computed on the same data indicated a one- or two-factor solution. Nguyen, Biderman, and McNary (2010) administered the CCAI to 175 undergraduate and MBA students at a U.S. university, along with the International Personality Inventory Pool (IPIP; available at http://ipip.ori.org/ipip/). CFA indicated that the four-factor structure did not fit the data well, and there were no correlations between the CCAI scales and the IPIP scales; evidence for validity was obtained only when the CCAI was augmented with items from the IPIP.

The only study to test the ecological validity of the CCAI was by Goldstein and Smith (1999), who compared 42 graduate student sojourners in the United States who had received intercultural training with a control group of 39 matched sojourners who did not. The training group had significantly higher scores than the control group and the norms reported by Kelley and Meyers (1987) on the total score and all four scale scores, but there was no pre-test administered.

Cross-Cultural Sensitivity Scale (CCSS)

The CCSS was designed to measure the "valuation and tolerance of different cultures" of dominant group members in Canada (Pruegger & Rogers, 1993). An initial 140-item pool was

generated based on the available cross-cultural psychology literature, a survey of Canadian attitudes toward immigrants, the researchers' experiences, and items from a measure of culture shock and racial attitudes and values. The items assessed cultural knowledge, attitudes, beliefs, and lifestyles. A three-member panel reviewed the items to eliminate redundancy, grammatical errors, double-barrelledness, and negative wording, resulting in a pool of 118 items. The item pool was initially administered to 55 undergraduate psychology students and 10 geologists, who also completed the Social Desirability subscale of the Personality Research Form (Jackson, 1974). A total score was calculated across the 118 items and items with nonsignificant item-total correlations were removed, resulting in 53 items. Items were then further removed based on their correlations with social desirability, resulting in a final selection of 24 items. Two parallel forms were then generated by rank ordering the items based on their item-total correlations and alternately selecting items to create two 12-item forms. There were no differences between the means on the two forms, and they were highly correlated with each other, $r(55) = .97$, and with the 24-item total scale, $r(55) = .97$. A second sample of 71 undergraduate students then completed the scales (αs = .87 and .80), and there was no difference between the means of the two scales.

The construct validity of the CCSS was tested in a study (Klein, 1995) involving 54 Canadian elementary schoolchildren Grades 3, 5, and 6, who were also administered a standard intelligence test (*Wechsler Intelligence Scale for Children, Third Edition*). CCSS scores were significantly correlated with Verbal IQ and Full Scale IQ. Ecological validity was tested in a study examining the efficacy of a semester-long elective course on multiculturalism involving 16 students, who completed the CCSS before and after the class (Klein, 1994); there were no significant differences between the pre- and posttest scores.

CQ

Ang, Van Dyne, and Koh (2006) defined CQ as an individual's capability to deal effectively in situations characterized by cultural diversity. They reviewed literatures on intelligence and intercultural competencies, educational and cognitive psychology operationalizations of metacognition, intrinsic satisfaction, self-efficacy, and intercultural communication, and conducted interviews with eight executives with extensive global work experience (Ang et al., 2007). They developed the CQ based on Earley and Ang's (2003) theoretical model that characterized CQ as comprising four components: metacognitive, cognitive, motivational, and behavioral CQ. Metacognitive CQ refers to the processes by which individuals acquire and understand cultural knowledge; cognitive CQ is general knowledge about culture; motivational CQ is the magnitude and direction of energy applied toward learning and functioning in cross-cultural situations; and behavioral CQ is the capability to exhibit appropriate actions when interacting with people of different cultures. Fifty-three items were initially drafted, which were ranked by three faculty and three international executives with cross-cultural expertise for clarity, readability, and definitional fidelity. The 10 best items were retained for each dimension (Ang et al., 2007). The 40-item CQ was given to 576 Singaporean undergrads. Items with high residuals, low factor loadings, small standard deviations, or extreme means and low item-total correlations were removed, resulting in a 20-item scale, with 4, 6, 5, and 5 items assessing metacognitive, cognitive, motivational, and behavioral CQ, respectively (αs = .70-.86).

A number of studies involving different groups of respondents have provided evidence for the construct validity of the CQ by confirming its four-factor structure using CFA. Samples have included Singaporean business undergrads (Ang et al., 2006), American and Singaporean undergrads (Ang et al., 2007, Study 1), a multicultural group of foreign professionals and their supervisors (Ang et al., 2007, Study 3), U.S. real estate agents (X.-P. Chen, Liu, & Portnoy, 2011), Pilipino laborers (A. S.-y. Chen, Lin, & Sawangpattanakul, 2011), organizational leaders and followers (Groves & Feyerherm, 2011), full-time employees (Imai & Gelfand, 2010), Korean undergrads (Moon, 2010), military leaders (Rockstuhl, Seiler, Ang, Van Dyne, & Annen, 2011),

and expatriates of Taiwanese manufacturing firms (Lee & Sukoco, 2010). These studies reported alphas for each of the scales above .70, often above .80. Additional evidence for the construct validity of the CQ comes from studies reporting correlations between CQ and personality traits (Ang et al., 2006; Ang et al., 2007; Fischer, 2011), emotional intelligence (Ang et al., 2007, Study 1; Groves & Feyerherm, 2011; Moon, 2010), leadership effectiveness (Rockstuhl et al., 2011), and cooperative negotiation heuristics (Imai & Gelfand, 2010). The CQ has also been correlated with the CCAI (Ang et al., 2007, Study 1) and the Openmindedness Scale from the MPQ (Fischer, 2011), providing evidence for its convergent validity.

There is considerable evidence for the concurrent and predictive ecological validity of the CQ with samples from multiple cultures. Scores on the CQ have predicted cross-cultural judgment and decision making, general and interactional adjustment and well-being, task performance on a problem-solving simulation, and work performance (Ang et al., 2007); cultural sales (i.e., the number of housing transactions occurring between people of different cultural origins; X.-P. Chen et al., 2011); culture shock and work performance (A. S.-y. Chen et al., 2011); organizational innovation and transformational leadership behaviors (Elenkov & Manev, 2009); leader and team performance (Groves & Feyerherm, 2011); cooperative relationship management behaviors (Imai & Gelfand, 2010); cultural adjustment (Lee & Sukoco, 2010; Templer, Tay, & Chandrasekar, 2006); travel stress (Ramsey, Leonel, Gomes, & Monteiro, 2011); and psychological adjustment and sociocultural adaptation (Ward, Wilson, & Fischer, 2011). Pre–post tests of the efficacy of intercultural training using the CQ as an outcome measure, however, have provided mixed results, with one report providing positive findings (Hodges et al., 2011) and one reporting negative findings (Fischer, 2011). Several studies have provided evidence for the incremental validity of the CQ to predict adjustment or adaptation above and beyond variables related to personality, demographics, and emotional intelligence (Ang et al., 2007, Studies 2 and 3; X.-P. Chen et al., 2011; Groves & Feyerherm, 2011; Imai & Gelfand, 2010).

IBA and BASIC

Ruben (1976) developed the IBA as a method to bridge the gap between intercultural knowledge and performance as a way to evaluate intercultural training efficacy. He contended that what a person knows is often not reflected well in behavior, and that adaptations in behaviors are what are important to 3C. Based on the literature on inter- and intracultural competence, Ruben identified seven aspects of behaviors he deemed important to an evaluation of 3C: display of respect, interaction posture, orientation to knowledge, empathy, self-oriented role behavior, interaction management, and tolerance for ambiguity. Self-oriented role behavior was further subdivided into three components, producing a nine-dimension scale. Operational definitions were created for each dimension so that each was characterized in terms of specific and observable behaviors. Each dimension was rated on either a 4- or 5-point scale, with behavioral descriptions associated with each anchor point. The test was intended to be easily administered by untrained observers with efficiency and reliability.

Interrater reliability was tested in an initial study of 19 individuals scheduled for assignments in Kenya at the conclusion of a 7-day intercultural adaptation training program and who were rated by three staff members (Ruben, 1976). All interrater correlations were statistically significant. Ratings were averaged across raters and the data were Q-factor analyzed to identify clusters of individuals. The analysis yielded three clusters of individuals.

The predictive ecological validity of the IBA was documented in a follow-up study of the 19 individuals (Ruben & Kealey, 1979). One year after arriving in Kenya, the participants were assessed on culture shock, four indicators of psychological adjustment, and three indicators of intercultural effectiveness. Culture shock was predicted by six dimensions assessed by the IBA, while psychological adjustment and intercultural effectiveness were each predicted by two and four dimensions, respectively.

Koester and Olebe (1988) revised the IBA to be completed by laypersons about roommates. Calling the new scale BASIC, they administered it to three groups of participants ($N = 263$) in a study of intercultural communication differences between inter- and intracultural roommates in a university residence hall about 10 weeks after the beginning of the semester (αs for the three groups = .77, .80, and .88). The total BASIC score was correlated with a single-item rating of communication effectiveness of the roommate at .60. An EFA on the BASIC produced a single factor that was labeled intercultural communication effectiveness.

Using the same data, Olebe and Koester (1989) tested the cross-cultural equivalence of the BASIC by computing EFAs separately for U.S. and non-U.S. students; both produced single-factor solutions. ANOVAs testing differences between the two groups on the individual BASIC dimensions and the total score produced no differences. Regressions of the BASIC scale scores on the single-item measure of communication effectiveness were significant for the U.S. and non-U.S. groups.

Graf and Harland (2005) administered the BASIC, Intercultural Sensitivity Scale (ISS; see below), and three measures of interpersonal competence to 188 MBA students at a Midwestern U.S. university. Participants also read an intercultural scenario and wrote decision solutions to it, which were rated independently by two MBA students for intercultural decision quality. Alpha for the BASIC total score was .59. The BASIC was not correlated with the three interpersonal measures or with the ISS. Two BASIC scales were significantly correlated with intercultural decision quality.

Intercultural Adjustment Potential Scale (ICAPS)

Matsumoto and colleagues (2001) developed the ICAPS as a way to assess the potential for intercultural adjustment as a function of the psychological skills that individuals possess, and identified eight constructs to assess: emotion regulation, critical thinking, openness, flexibility, interpersonal security, emotional commitment to traditional ways of thinking, tolerance for ambiguity, and empathy. An initial item pool of 193 items was created. Item reduction was achieved first through an ecological validity test with 28 Japanese sojourners who self-rated their adjustment, length of stay in the United States, and academic performance (Matsumoto et al., 2001, Study 1). Items not correlating with any criterion variable were eliminated, resulting in 153 items. A second ecological validity test involving focus group discussions about adjustment to the United States with a different sample of 34 Japanese sojourners was conducted (Matsumoto et al., 2001, Study 2). Criterion variables included self-, peer-, and facilitator ratings of adjustment based on the focus group discussions, and the same criterion variables in Study 1. Items that had the lowest summed p values across all criterion variables were kept, along with items that were correlated significantly with any of the criterion variables. This resulted in a 55-item ICAPS ($\alpha = .78$). Test–retest and parallel forms reliability were tested in 25 respondents from the initial studies who completed English and Japanese versions of the ICAPS 1 to 2 months after initial administration (Matsumoto et al., 2001, Study 3). Test–retest correlations were .79 and .84 for both versions, respectively; parallel forms reliability was .93. Parallel forms reliability with a Spanish version of the test was also subsequently documented (Matsumoto et al., 2003, Study 6).

A number of studies have provided support for the concurrent ecological validity of the ICAPS with Japanese student sojourners and immigrants (Matsumoto et al., 2001, Study 4), Japanese nonstudent sojourners and immigrants (Matsumoto et al., 2003, Study 1), and Japanese females in intercultural marriages (Matsumoto et al., 2003, Study 3). The predictive ecological validity of the ICAPS was established in Japanese sojourners who completed the ICAPS in their home country 1 month prior to their sojourn and later completed measures of culture shock, homesickness, satisfaction with life, and subjective adjustment in the host country (Matsumoto et al., 2003, Study 4). Pre–post ecological validity was further established in two studies involving Japanese

sojourners who participated in an intercultural seminar (Matsumoto et al., 2003, Study 2; Matsumoto et al., 2001, Study 6); posttest scores on the ICAPS were significantly higher than pretest scores measured 1 month before the sojourn.

Although the ICAPS was originally developed for use with Japanese immigrants and sojourners, subsequent studies documented its concurrent ecological validity with U.S. undergrads (Matsumoto, LeRoux, Bernhard, & Gray, 2004, Studies 1 and 2; Matsumoto et al., 2001, Study 5); immigrants and sojourners to the United States, including samples from India, Sweden, and Central and South America (Matsumoto et al., 2003, Study 5; Matsumoto, LeRoux, Robles, & Campos, 2007; Yoo, Matsumoto, & LeRoux, 2006); and Spanish-speaking sojourners to the United States (Matsumoto et al., 2003, Study 6). Ecological validity using extreme group comparison was documented in a study involving intercultural counselors and consultants, who also provided ratings of adjustment and years in occupation (Matsumoto et al., 2001, Study 8). Ecological validity was also demonstrated using a test of behaviors (an In-Basket task), and incremental validity of the ICAPS to predict behavioral outcomes over and above emotion recognition abilities was demonstrated in the same study (Matsumoto et al., 2004, Study 3). Predictive ecological validity was documented in a sample of international students in the United States 2 and 9 months after initial ICAPS assessment (Yoo et al., 2006), and in a sample of U.S. undergrads who studied abroad (Savicki, Downing-Burnette, Heller, Binder, & Suntinger, 2004).

Three of the studies reported above (Matsumoto et al., 2004, Studies 1 and 2; Matsumoto et al., 2001, Study 5) provided data supporting the incremental and discriminant ecological validity of the ICAPS to predict adjustment above and beyond personality measures and the CCAI. ICAPS predicted future adjustment even after controlling for emotion recognition ability (Yoo et al., 2006). The ICAPS has also predicted anxiety, hopelessness, satisfaction with life, culture shock, and contentment above and beyond personality and intelligence as measured by the Wechsler Adult Intelligence Scale (Matsumoto et al., 2007); in the same study, intelligence did *not* predict adjustment above and beyond what was already predicted by ICAPS and personality.

Evidence for the construct validity of the ICAPS comes from an initial EFA of the 55-item ICAPS involving 1,751 respondents, which generated a four-factor solution that accounted for 18.6% of the cumulative variance in the data. The four factors were labeled Emotion Regulation, Openness, Flexibility, and Critical Thinking. Additional evidence for construct validity was reported in two studies documenting correlations between the ICAPS, the CCAI, and a big five personality measure (Matsumoto et al., 2004, Study 1; Matsumoto et al., 2001, Study 5); with the California Psychological Inventory, a measure of altruism, and the Myers-Briggs Typology Inventory (Matsumoto et al., 2004, Study 2); and with a big five personality measure, optimism–pessimism, hope, and coping (Savicki et al., 2004).

Intercultural Communication Competence (ICC)

The ICC was developed based on interviews with 15 student and nonstudent volunteers with experience in intercultural communication, and that resulted in the identification of five characteristics associated with ICC: empathy, intercultural experience and training, motivation, global attitude, and ability to listen well in conversation (Arasaratnam & Doerfel, 2005). These were summarized into three dimensions—cognitive, affective, and behavioral—and five items were generated to assess each dimension, resulting in a 15-item test (Arasaratnam, 2009). Items in the cognitive dimension were inspired by research on cognitive complexity in relation to communication competence. Items in the affective dimension were related to empathy, while items in the behavioral dimension were associated with interaction seeking, adapting behaviors or communication patterns according to the other, and engaging in friendships with people from other cultures.

Evidence for construct validity was provided in a study involving 174 Australian undergraduates and 127 international students in Sydney (Arasaratnam, 2009). The ICC was administered along with measures of attitudes toward other cultures, ethnocentrism, motivation, and interaction involvement. EFA on the 15-item ICC generated a single-factor solution. A total score was created ($\alpha = .77$) and all four measures were significantly correlated with the ICC total. Additional evidence for construct validity came in a subsequent study (Arasaratnam & Banerjee, 2011) involving 125 Australian undergraduates and 106 international students in Sydney who completed the ICC and measures of ethnocentrism, sensation seeking, motivation to engage in intercultural communication, and attitudes toward other cultures. Ethnocentrism, motivation, and attitudes were once again significantly correlated with ICC; sensation seeking was negatively correlated with ethnocentrism and positively correlated with motivation to engage in intercultural communication.

ICSI

Bhawuk and Brislin (1992) developed the ICSI based on the concept of intercultural sensitivity, defined as sensitivity to the importance of cultural differences and to the points of view of people of other cultures. It was designed to measure the ability of people to modify their behavior in intercultural contexts by examining people's understanding of the different ways they can behave depending upon whether they are interacting in an individualistic or a collectivist culture, their open-mindedness concerning the differences they encounter in other cultures, and their flexibility concerning behaving in unfamiliar ways that are called upon by the norms of other cultures. Items were drafted based on the available literature on adjustment and adaptation, and previous work involving the analysis of critical incidents and cross-cultural training (Brislin, Cushner, Cherie, & Yong, 1986; Triandis, Brislin, & Hui, 1988). The initial item pool included 71 items, 26 assessing individualism versus collectivism (rated twice = 52 items), 10 assessing open-mindedness, and 9 assessing flexibility. All items were drafted to capture behaviors rather than attitudes or traits. Initial testing involved 46 MBA students and 93 graduate students at the East–West Center at the University of Hawaii, many of whom were international student professionals (Bhawuk & Brislin, 1992). Items with item-total correlations > .1 were retained, which resulted in a final version of the instrument with 46 items, 16 individualism versus collectivism items rated twice and 14 items assessing flexibility and open-mindedness. Alphas for the two samples were .82 and .84, respectively.

Bhawuk and Brislin (1992) provided initial evidence for construct validity by conducting EFAs on the two sets of items, which indicated that the first set of items factored into two scales consistent with individualism versus collectivism. The EFA on the second set of items indicated that 9 and 11 of the 14 items loaded on the intended scales for the two samples. Additional evidence for construct validity was reported by Kiuchi (2006), who administered the ICSI and a measure of self-efficacy to 85 Japanese college students in the United States. The self-efficacy measure produced two scale scores, one for general self-efficacy, the other for social efficacy (Scherer et al., 1982). As predicted, the ICSI was correlated with general self-efficacy but not with social efficacy.

Bhawuk and Brislin (1992) examined the ecological validity of the ICSI by testing the differences on the ICSI total score among various demographic variables. Individuals who reported trying a greater number of ethnic goods, number of years spent abroad, and an interest in working with students from other cultures all had higher scores on the ICSI than those who did not. There were no differences in ICSI scores on the number of friends from other cultures or the number of languages spoken. Also, nine program staff members from the East–West Center rated a subsample of the students on four items assessing their intercultural interactions. Mean ratings on the items were dichotomized into high and low groups and their difference was significant.

Ecological validity was further tested by Bhawuk (1998) in a study involving 102 exchange students at a Midwestern U.S. university. Participants were assigned to one of four conditions, three of which involved cross-cultural training involving culture-specific, culture-general, or culture theory–based cultural assimilators, and the fourth being a control group that received a reading assignment. Two of the three training groups had significantly higher scores on the ICSI than the control group.

IDI

Hammer, Bennett, and Wiseman (2003) developed the IDI based on Bennett's (1986) Developmental Model of Intercultural Sensitivity, which proposes six stages of development that reflect how an individual moves from a stage of ethnocentrism to ethnorelativism, labeled Denial, Defense Renewal, Minimization, Acceptance, Adaptation, and Integration. Hammer and colleagues developed the IDI by first interviewing 40 international student volunteers in the United States, reviewing a randomly selected sample of 25 transcripts, and rating the interviewee's orientations according to the six stages. The researchers then selected statements from the 40 interviews that exemplified each stage. After eliminating redundant items, the authors created new items for some stages and produced an initial item pool of 239 items. A panel of intercultural experts reviewed the item pool; items that could not be reliably categorized by an agreement rating of .60 or larger were eliminated, resulting in a 145-item pool. The 145-item IDI was administered to 226 participants. Instead of conducting a single EFA on all items, the authors computed six separate EFAs on the items intended for the six scales. They then kept items on each scale that loaded at least .50 with a cross-loading < .20. This resulted in a 60-item IDI, 10 items on each of six scales (αs ranging from .80 to .91). The six scales, however, did not match the six stages of Bennett's model; the Reversal and Integration stages were not reproduced and the Adaptation stage was split into two scales. Paige, Jacobs-Cassuto, Yershova, and DeJaeghere (1999; cited in Hammer et al., 2003) administered the 60-item IDI to 330 respondents and conducted an EFA on all 60 items. This also did not produce factors that corresponded to Bennett's six stages on which the IDI was based. Thus, Hammer and colleagues reverted back to the 145-item IDI. By editing and selecting alternative items, they then produced a 122-item IDI. They administered this scale to 591 college students, along with scales measuring worldmindedness and intercultural anxiety. CFA indicated that a five-scale model best fit the data. Scale scores were computed (αs = .80-.85). Worldmindedness and intercultural anxiety correlated with three of the five scales, providing some evidence for construct and ecological validity.

Subsequent studies continued to produce mixed results concerning the construct validity of the IDI. Paige, Jacobs-Cassuto, and Yershova (2003) administered the 60-item IDI to 353 college and high school students (it is not clear whether this was the same sample as reported in Hammer et al., 2003); once again EFA on the 60 items did not support the six-stage model. Greenholtz (2005) administered the IDI to 400 respondents and PCA generated a seven-factor solution, with 27% of the items not loading on the intended factors. Hammer (2011), however, administered a 50-item IDI to 4,763 respondents from 11 cultural groups, and CFA produced a seven-factor model that roughly corresponded to the Bennett (1986) model. Interscalar correlations were loosely in the predicted direction.

A few studies have provided evidence of the ecological validity of the IDI with demographic variables such as intercultural experience, friends from other cultures, and language study (Paige et al., 2003); age, father's education, and years spent in another culture (Yuen, 2010); and length of time attending international school (Straffon, 2003). Pre–post tests using the IDI to assess training efficacy, however, have provided mixed results, with some studies providing positive results (Anderson, Lawton, Rexeisen, & Hubbard, 2006; DeJaeghere & Cao, 2009; Hammer, 2011) and some negative (Atshuler, Sussman, & Kachur, 2003; Pedersen, 2010). Two case

studies using the IDI have been published (Cassiday, 2005; Greenholtz & Kim, 2009), and the IDI has been modestly correlated with conflict styles in a sample of teachers (Mahon, 2009).

ISS

G.-M. Chen and Starosta (2000) focused on the affective aspects of intercultural communication in the development of the ISS, drafting 73 items that assessed six affective elements proposed to be important to intercultural sensitivity: self-esteem, self-monitoring, open-mindedness, empathy, interaction involvement, and suspending judgment. In a pilot study, G.-M. Chen and Starosta administered the initial item pool to 168 1st-year students in basic courses in communication studies and kept 44 items with item-total correlations > .50. They administered the 44 items to 414 college students and conducted an EFA, which generated five factors accounting for 37.3% of the variance, labeled Engagement, Respect for Cultural Differences, Interaction Confidence, Interaction Enjoyment, and Interaction Awareness (G.-M. Chen & Starosta, 2000, Study 1). They retained 24 items with item loadings ≥.50 with no cross-loadings ≥.30. G.-M. Chen and Starosta (Study 2) then administered the 24-item ISS and scales measuring interaction attentiveness, impression rewarding, self-esteem, self-monitoring, and perspective taking to 162 undergraduates. Alpha was .86 and significant correlations were obtained with all five measures. G.-M. Chen and Starosta (Study 3) provided initial evidence for the ecological validity of the ISS by administering it and scales measuring intercultural effectiveness and attitudes toward intercultural communication to 174 undergraduates. Significant correlations were obtained between the ISS and both scales.

As described earlier, Graf and Harland (2005) administered the ISS, BASIC, and three measures of interpersonal competence to 188 MBA students at a Midwestern U.S. university. Participants also read an intercultural scenario and wrote decision solutions to it, which were rated independently by two MBA students for intercultural decision quality. Alphas for the six ISS scales and total score ranged from .47 to .89. The ISS was not correlated with the three interpersonal measures or the BASIC. Four ISS scales were correlated with intercultural decision quality; in addition, a regression including all intercultural and interpersonal scales indicated that ISS total scores were predictive of intercultural decision quality, supporting its ecological validity.

MPQ

van der Zee and van Oudenhoven (2000) defined multicultural effectiveness as successfully operating in a new cultural environment, a feeling of psychological well-being in that environment, and interest in and ability to deal with individuals from a different cultural background. They designed the MPQ by first identifying personality dimensions they considered on an a priori basis to predict this definition of multicultural effectiveness. Based on their review of the literature, the authors considered the dimensions Cultural Empathy, Openmindedness, Emotional Stability, Orientation to Action, Adventurousness/Curiosity, Flexibility, and Extraversion. For each MPQ dimension, the authors generated items that described concrete behaviors or tendencies they considered indicative of the specific dimension. The original scale included a total of 91 items—14, 13, 13, 12, 12, 12, and 15 items—assessing each dimension, respectively.

Initial construct validation involved an EFA that produced a four-factor solution explaining 30.6% of the total variance in the data set (van der Zee & van Oudenhoven, 2000). Some of the original scales were therefore collapsed, resulting in final factors labeled Openness, Emotional Stability, Social Initiatives, and Flexibility. Alphas ranged from .75 to .90, and test–retest correlations ranged from .75 to .87. Construct validity was supported by moderate levels of interscalar correlations, and with significant correlations with the big five personality traits and measures of

need for change and rigidity. Ecological validity was initially demonstrated by significant correlations with variables assessing multicultural activities, aspirations for an intercultural career, and intercultural orientation. The incremental ecological validity of the MPQ was demonstrated by hierarchical regressions that indicated that some MPQ scales predicted these criterion variables even when big five personality traits were controlled.

Subsequent studies involving the MPQ utilized a five-factor scoring procedure, with scales labeled Emotional Stability, Social Initiative, Openmindedness, Cultural Empathy, and Flexibility and items varying from 77 to 91. Using different versions, a number of subsequent studies have provided evidence for the reliabilities of the scale scores (αs between .64 and .92; Bakker, van der Zee, & van Oudenhoven, 2006; Bobowik, van Oudenhoven, Basabe, Telletxea, & Paez, 2011; Herfst, van Oudenhoven, & Timmerman, 2008; Houtz, Ponterotto, Burger, & Marino, 2010; Korzilius, van Hooft, Planken, & Hendrix, 2011; Leong, 2007; van der Zee & van Oudenhoven, 2001; van der Zee, van Oudenhoven, & de Grijs, 2004; van der Zee, Zaal, & Piekstra, 2003; van Oudenhoven, Mol, & van der Zee, 2003; van Oudenhoven & van der Zee, 2002). At least two studies have reported alphas for an overall MPQ score, presumably using the sum of all items (Simkhovych, 2009; Ward, Fischer, Lam, & Hall, 2009).

The construct validity of the MPQ has been demonstrated in a number of ways in various studies. Many studies have reported intercorrelations among the scale scores in the predicted fashion (Houtz et al., 2010; Korzilius et al., 2011; Leone, van der Zee, van Oudenhoven, Perugini, & Ercolani, 2005; van der Zee et al., 2003; van der Zee et al., 2004; van der Zee & van Oudenhoven, 2001; van Oudenhoven et al., 2003; Ward et al., 2009). Studies involving different samples using PCA or CFA have replicated the five-factor structure of the MPQ (van der Zee et al., 2003), and at least one study has confirmed the factor structure of the MPQ in two different cultural groups (Italians and Dutch; see Leone et al., 2005). (One study—Ponterotto et al., 2007—however, reported a three-factor solution.) The MPQ scales have correlated significantly with a number of other personality constructs, including self- and other ratings (van der Zee & van Oudenhoven, 2001); big five personality traits (Leone et al., 2005; van der Zee et al., 2003); verbal abilities, vocational interests, and occupational group (van der Zee et al., 2003); growth-oriented values (Bobowik et al., 2011); problem solving (Houtz et al., 2010); emotional intelligence (Ponterotto, Ruckdeschel, Joseph, Tennenbaum, & Bruno, 2011); family cohesion, adaptability, and communication (Ali, van der Zee, & Sanders, 2003); and the 20-item CQ measure (Ward et al., 2009).

A number of studies have provided evidence for the ecological validity of the MPQ by reporting significant relationships between it and international career aspirations, multicultural activities, aspired professions, international orientation, and self-ratings and aptitude for an international career (Leone et al., 2005; van der Zee & van Oudenhoven, 2000, 2001); number of foreign languages spoken and self-rated proficiency (Korzilius et al., 2011); and experience living abroad (van der Zee & van Oudenhoven, 2001). Other studies have documented the ecological validity of the MPQ in predicting more traditional and direct measures of adaptation and adjustment, including satisfaction with life, physical and mental health, social interactions, and academic achievement (van Oudenhoven et al., 2003; van Oudenhoven & van der Zee, 2002); job satisfaction (van Oudenhoven et al., 2003); team commitment and exam performance in groupwork (van der Zee, Atsma, & Brodbeck, 2004); anxiety, positive and negative affect, and appraisals (van der Zee et al., 2004); reactions to hypothetical scenarios concerning acculturation strategies (Bakker et al., 2006); team identification and affect in diverse teams (van der Zee & van der Gang, 2007); responses to critical incidents and self-ratings of intercultural experiences (Herfst et al., 2008); sociocultural adaptation and depression (Leong, 2007; Ward et al., 2009); academic performance, experienced difficulties, social support, psychological health, and satisfaction with life (Long, Yan, & van Oudenhoven, 2009); stress and homesickness (Suanet & van de Vijver, 2009); life satisfaction, intercultural interaction, and sociocultural adjustment (Ali et al., 2003); and subjective well-being (Ponterotto et al., 2007). There is also evidence for the ecological validity of the

MPQ using behavioral measures (van der Zee et al., 2003) and for its predictive ecological validity (van Oudenhoven & van der Zee, 2002).

A number of the studies cited immediately above have documented the incremental ecological validity of the MPQ to predict successful adjustment outcomes above and beyond other personality measures (van der Zee et al., 2003; van der Zee & van Oudenhoven, 2000), self-efficacy (van Oudenhoven & van der Zee, 2002), demographics (van Oudenhoven et al., 2003), and CQ (Ward et al., 2009). Several studies have also provided evidence for ecological validity using extreme group difference tests comparing international studies students and psychology students (van der Zee & van Oudenhoven, 2001), third culture kids and single language kids (Dewaele & van Oudenhoven, 2009), students in an exchange program versus undergraduate controls (Leong, 2007), and international versus noninternational business employees (Korzilius et al., 2011).

Evaluation of the Available 3C Tests

Limitations

Before evaluating the evidence presented above, we acknowledge several limitations of our approach, and readers should interpret our descriptions above and our evaluations below with these caveats. First, we relied solely on empirical articles published in peer-reviewed journals because the peer-review process allows for a quality control check on the information that is deposited into the literature, and authors need to provide enough methodological detail for readers to evaluate the quality of the findings reported. Books, chapters, theoretical or review articles, and unpublished manuscripts certainly may include information about the psychometric properties of tests not found in peer-reviewed journal articles, but they do not pass the quality control process of peer review. Also much of the information presented in such publications overlaps with that reported in peer-reviewed articles, which would have been problematic to disentangle.

We also did not include government or private industry reports because these also do not provide the quality control of the peer-review process. Also many are not available in open sources and are limited to government access or are proprietary information for commercial businesses. Thus, any evidence used from such sources would be arbitrary and could not involve a complete review of all data sources.

Finally, we utilized only articles published in English. Articles reporting validity and reliability efforts of 3C tests are available in other languages (e.g., Bakker, van der Zee, & van Oudenhoven, 2003; Leone, Lucidi, Ercolani, & Presaghi, 2003; Yan, 2009). One hesitation with including them is that we cannot be sure that the data reported in the non-English reports are also not reported in English reports. Also it is impossible to survey all reports in all non-English languages, and the inability to do so would result in an unbalanced review of the literature. Thus, for consistency, we limited the search to articles published in English.

Evaluation

Content validity. One of the criteria we described earlier with which to evaluate the content validity of the available 3C tests concerned whether the initial item pools exhausted the possible universe of measurement for the KSAOs identified. Two tests—ICC and IBA/BASIC—are questionable in this regard; the former included only 15 items to assess three broad constructs, and the IBA/BASIC included only one item per identified construct. The content validity for the remaining tests was adequate and appropriate for the definition of 3C they operationalized and the KSAO domains identified. There were, however, major differences across the tests in those definitions and domains. The CCAI, CCSS, CQ, ICAPS, and MPQ considered KSAOs necessary for general intercultural adaptation and adjustment. The ICC and ISS focused on intercultural communication sensitivity, the ICSI assessed intercultural sensitivity specific to individualism

and collectivism, and the IDI assessed a specific theoretical model of the development of intercultural sensitivity. Thus, understanding differences among 3C tests requires acknowledgment of the specific domain(s) of 3C they intended to assess in the first place.

With regard to validation approach, the CCAI, CQ, IBA/BASIC, ICC, ICSI, MPQ, IDI, and ISS utilized the construct validity–driven approach, which involved first the identification of the latent structure underlying an initial item pool and item reduction through the removal of items not associated to the latent structure. The ICAPS utilized the ecological validity–driven approach, which involved first testing the ecological validity of individual items of the item pool and item reduction by removing items not associated with criterion variables. Evaluation of the construct and ecological validity evidence for these tests, therefore, should be interpreted vis-à-vis the advantages and disadvantages of these approaches described earlier. We could not find clear information for the CCSS that would elucidate its validation approach. An initial item pool of 178 items was reduced to 24 based solely on item-total correlations. This item reduction procedure assumes a single factor underlies the initial 178-item scale, which is highly unlikely and antithetical to the test's own content validation model.

Construct validity. Our review indicates that the evidence for the structural validity of a number of 3C tests is lacking. For example, although the CCAI was designed to assess four constructs, attempts to validate its four-factor structure have not been successful (Davis & Finney, 2006; Nguyen et al., 2010). We found no study that assessed the factor structure of the CCSS or the ICC. The factor structures of the IBA/BASIC, the ICSI, and the ISS were only assessed in a single U.S. and/or international student sample (Bhawuk & Brislin, 1992; G.-M. Chen & Starosta, 2000; Olebe & Koester, 1989). And assessments of the factor structure of the IDI have produced inconsistent results in limited samples that do not correspond to the model intended to be measured (Greenholtz, 2005; Hammer, 2011; Hammer et al., 2003; Paige et al., 2003). Nonconfirmation of structural validity renders correlations of scale scores to other constructs questionable because it is not clear what the scales are assessing.

The construct validity of the ICAPS is marginally adequate. On one hand, EFA involving student and nonstudent respondents from multiple cultures produced a four-factor solution that corresponded to some of the constructs proposed related to 3C. The factor structure, however, only accounted for 18.6% of the total variance in the data set and generated low alphas (.43-.64). On the other hand, the reliability of the ICAPS total score, which is not a mathematical derivative of the scale scores, was acceptable, as were its test–retest and parallel forms reliabilities; ICAPS scores have been correlated with personality scales and with at least one other test of 3C (CCAI), and correlations with personality survive after controlling for the CCAI. ICAPS scores have also been correlated with an ability test of emotion recognition in two studies and with a test of general intelligence.

The construct validity of the CQ and MPQ is strong. For the CQ, the four-factor structure has been confirmed in multiple studies with student and nonstudent respondents from multiple cultures, including the United States, Singapore, South Korea, Taiwan, and others. High scale reliabilities have consistently been reported, and CQ has been reliably correlated with personality traits, emotional intelligence, leadership style, and other tests of 3C. For the MPQ, factor analyses have generally supported its structure in multiple studies in different cultures with student and nonstudent samples. Many studies have reported intercorrelations among the scale scores in the predicted directions and high alphas. Many of these same studies have documented correlations between the MPQ and measures of personality, intelligence, vocational interests and occupational group, values, problem-solving styles, and at least one other test of 3C (CQ).

Ecological validity. As mentioned above, we believe that the main criterion against which a test of 3C should be evaluated is the strength of the evidence for its ecological validity. We summarized the evidence for ecological validity for all ten 3C tests in Table 1. Instead of imposing

Table 1. Summary of the Available Evidence for the Ecological Validity of the Ten 3C Tests.

Test	Used valid and reliable criterion variables	Positive effects from pre–post sojourn or training	Extreme group comparison	Concurrent ecological validity	Predictive ecological validity	Incremental validity over demographics	Incremental validity over personality	Incremental validity over at least one other 3C test	Cross-cultural samples (beyond international sojourners in the United States)	Mixed methodologies
CCAI	No	No	No	No	No	No	No	No	No	No
CCSS	No	No	No	No	No	No	No	No	No	No
CQ	Yes	Mixed	No	Yes	Yes	Yes	Yes	No	Yes	Yes
IBA/BASIC	No	No	No	Mixed	Yes	No	No	No	No	No
ICAPS	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ICC	No	No	No	No	No	No	No	No	No	No
ICSI	No	Mixed	No	Mixed	No	No	No	No	No	No
IDI	No	Mixed	No	Yes	No	No	No	No	No	No
ISS	Yes	No	No	Yes	No	No	No	No	No	No
MPQ	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note. 3C = cross-cultural competence; CCAI = Cross-Cultural Adaptability Inventory; CCSS = Cross-Cultural Sensitivity Scale; CQ = Cultural Intelligence Scale; IBA = Intercultural Behavioral Assessment; BASIC = Behavioral Assessment Scale for Intercultural Communication Effectiveness; ICAPS = Intercultural Adjustment Potential Scale; ICC = Intercultural Communication Competence; ICSI = Intercultural Sensitivity Inventory; IDI = Intercultural Development Inventory; ISS = Intercultural Sensitivity Scale; MPQ = Multicultural Personality Inventory.

our subjective evaluation of the quality of the ecological validity data presented, the entries in the table refer to whether evidence was available for each of the types of ecological validity we considered. That is, if there were *any* data supporting a particular type of ecological validity, we entered "yes." If there were none, we entered "no." When positive and null findings have been reported, we entered "mixed." The columns are independent of each other; for example, a "yes" in cross-cultural samples meant that there was at least one study that involved cross-cultural samples beyond student sojourners to the United States, not that all studies did so. Moreover, not all entries are equivalent; a *yes* was entered if there was at least one study that provided evidence for each type of ecological validity, whether there was only 1 or 10 studies.

Table 1 makes it clear that the evidence for the ecological validity of a number of 3C tests is inadequate because they lack the use of valid and reliable criterion variables of intercultural adjustment or adaptation, have limited breadth of cross-cultural samples, do not use mixed methodologies, and do not provide evidence of concurrent, predictive, and/or incremental ecological validity. For example, there is only one study examining the ecological validity of the CCAI in a pre–post test of training efficacy in 42 international graduate students in the United States (Goldstein & Smith, 1999); although the training group had significantly higher posttest scores than the control group or norms, no pre-test was administered, leaving open the possibility that the differences were not due to training or assessment. Likewise, there was only one study testing the ecological validity of the CCSS, testing the efficacy of training on multiculturalism in 16 students (Klein, 1994); there were no significant differences between pre- and posttest scores. There were three studies providing evidence for the ecological validity of the IBA/BASIC; one, however, examined the predictive ecological validity of only 19 individuals posted to Kenya (Ruben & Kealey, 1979), and the other two only compared U.S. and non-U.S. undergrads on a single-item measure of communication effectiveness (Koester & Olebe, 1988; Olebe & Koester, 1989). No study has reported ecological validity of the ICC. For the ICSI, the only evidence for ecological validity comes from a single study reporting correlations with it and demographic variables and ratings of intercultural effectiveness in international students in the United States (Bhawuk & Brislin, 1992), and another study reporting pre–post training differences in university students in the U.S. Midwest (Bhawuk, 1998). The only evidence for the ecological validity of the IDI comes from studies reporting significant correlations between it and demographic variables (Paige et al., 1999; Straffon, 2003; Yuen, 2010); pre–post tests of training efficacy have provided mixed results (Anderson et al., 2006; Atshuler et al., 2003; DeJaeghere & Cao, 2009; Hammer, 2011; Pedersen, 2010). The ecological validity of the ISS is supported in only two studies, one reporting correlations with attitudes toward intercultural communication (G.-M. Chen & Starosta, 2000), and one reporting correlations with intercultural decision quality (Graf & Harland, 2005), both with only U.S. students.

The ecological validity of the CQ, ICAPS, and MPQ is strong. The CQ has predicted cross-cultural judgment and decision making, general and interactional adjustment and well-being, task performance on a problem-solving simulation, work performance, cultural sales, culture shock, organizational innovation and transformational leadership behaviors, leader and team performance, cooperative relationship management behaviors, cultural adjustment, travel stress, psychological adjustment, and sociocultural adaptation. Although there are some mixed findings using pre–post tests of the efficacy of intercultural training using the CQ as an outcome measure, several studies have provided evidence for the incremental validity of the CQ to predict adjustment or adaptation above and beyond variables related to personality, demographics, and emotional intelligence.

The ICAPS has predicted demographic variables associated with adjustment and adaptation, standardized tests of adjustment and adaptation, behavioral tasks, pre–post training efficacy, and

extreme group differences. The participant cultures have spanned the United States, Japan, Sweden, India, Central and South America, and others, and have included student and nonstudent samples. Ecological validity tests have involved mixed methodologies going beyond standard self-report scales, including self-, peer-, and facilitator ratings of behaviors in focus group interviews, behavioral tasks, participation in intercultural training seminars, and actual sojourns. Concurrent and predictive ecological validities have been demonstrated, along with incremental ecological validity of the ICAPS in predicting adjustment and adaptation above and beyond different demographic variables, personality measures, emotion recognition ability, standard intelligence, and at least one other 3C test (CCAI).

The MPQ has predicted demographic variables associated with adjustment and adaptation, standardized tests of adjustment and adaptation, behavioral tasks, and extreme group differences. The participant cultures have spanned the Netherlands, Belgium, the United States, the United Kingdom, France, Australia, Germany, Italy, Singapore, China, New Zealand, and others. The ecological validity tests used mixed methodologies beyond self-report, including behavioral tasks and interviews. Predictive and concurrent ecological validity, as well as the incremental ecological validity in predicting adjustment and adaptation above and beyond different demographic variables, personality measures, and at least one other 3C test (CQ) have been demonstrated. We did not, however, find evidence for pre–post tests of training efficacy; and one caution concerning the MPQ is that across reports, different scales have predicted different outcome variables, and some reports have also utilized a total score for the MPQ, which seems antithetical to its intended design and factor structure. Regardless, whenever any of the scores derived from the MPQ were associated with adjustment or adaptation, we liberally considered that as evidence for its ecological validity as a whole, but examination of specific scales may lead to different evaluations.

Conclusions and Recommendations

We conclude that the CQ, ICAPS, and MPQ have the most promising evidence for assessing 3C. At the same time, our review also indicates several important challenges for 3C tests to address in the future. For example, while a number of studies has confirmed the factor structure of the MPQ (albeit with some discrepancies), one recent study has suggested the existence of a single, general factor of personality that underlies many personality scales, and demonstrated that a single factor accounted for 41% of the reliable variance on four of the five MPQ scales (Rushton & Irwing, 2009). It is very possible that the same applies to the CQ and ICAPS as well. Future research examining the best-fit structure underlying the tests may address this concern.

Relatedly, there is a need for further psychometric testing of the various 3C tests. Although the CQ and MPQ have been subjected to multiple, independent CFAs, further research including structural equation modeling and CFA to test the goodness of fit of identified structures across cross-cultural samples of different demographics—sex, age, language, and so on—for all three tests is necessary. Except for the ICAPS, there appears to be a lack of evidence for test–retest reliability and parallel forms reliability in different languages, which is especially important for culture-general 3C tests (and guidelines for the interpretation of alpha reliabilities should be adopted; see Ponterotto & Ruckdeschel, 2007). While the CQ, ICAPS, and MPQ all report data involving mixed methodologies and incremental ecological validity above and beyond personality and other individual difference variables, certainly more research demonstrating incremental validity is necessary.

As mentioned previously, the suitability of any test is relative to the specific domain of 3C that it was intended to predict and the KSAOs that were theoretically hypothesized or empirically

Table 2. Four Domains That Comprise the Potential "Active Ingredients" of 3C.

Domain	CQ	MPQ	ICAPS
1	Motivation	Openmindedness; social initiative	Openness
2	Behavior	Openmindedness; flexibility	Openness; flexibility
3	Metacognition	Cultural empathy	Critical thinking
4		Emotional stability	Emotion regulation

Note. 3C = cross-cultural competence; CQ = Cultural Intelligence Scale; MPQ = Multicultural Personality Inventory; ICAPS = Intercultural Adjustment Potential Scale.

demonstrated to underlie the test. To wit, the CQ, ICAPS, and MPQ were all validated with some of the same outcome variables (e.g., stress, well-being). But each was also validated against unique variables not assessed by the others. For instance, the CQ predicted cognitive decision-making processes and leadership behaviors. The ICAPS predicted psychological adjustment, such as culture shock, depression, and anxiety. The MPQ predicted international and intercultural vocational interests and international orientation. Thus, different tests may differentially predict different outcomes. Future research involving a "World Cup" of these tests may explore this possibility further.

Examination of the latent structures of the CQ, ICAPS, and MPQ suggests an interesting list of the potential "active ingredients" that comprise 3C. There is conceptual overlap among the constructs assessed by all three scales, which suggests four major domains of 3C (Table 2). Domain 1 consists of CQ Motivation, MPQ Openmindedness, MPQ Social Initiative, and ICAPS Openness. Domain 2 consists of CQ Behavior, MPQ Openmindedness, MPQ Flexibility, ICAPS Openness, and ICAPS Flexibility. Domain 3 consists of CQ Metacognition, MPQ Cultural Empathy, and ICAPS Critical Thinking. Domain 4 consists of MPQ Emotional Stability and ICAPS Emotion Regulation. One study has demonstrated correlations between CQ and MPQ scales generally consistent with these groupings (Ward et al., 2009). The domains suggested in Table 2 are entirely speculative until verified, or not, by empirical studies. Future research examining the structure of the three tests combined should provide an empirical basis to identify the active ingredients of 3C. It is entirely possible that the resulting structures that emerge from such an analysis do not converge with the groupings presented in Table 2; we present them here as suggestions to encourage such work in the future.

One type of methodology that has been sorely missing in this literature and that may aid in the search for the active ingredients of 3C is that of qualitative research. Indeed, the development and validation of 3C tests can appear to be a process that is exclusive to quantitative methods. This does not have to be. Although case studies and interviews have been used in nearly all 3C test creation processes, they generally have been abandoned in the validation process (note, however, two case studies using the IDI). The incorporation of qualitative methodologies may add important flavor to the development of 3C tests that may complement the existing literature in important ways, and we encourage researchers to consider the inclusion of qualitative methods in future efforts (see review and discussion by Ponterotto, 2010).

Finally, incremental ecological validity tests of the CQ, ICAPS, and MPQ have demonstrated that each of these can predict cross-cultural outcomes above and beyond standard personality and intelligence. But although a number of studies have shown that CQ and ICAPS can predict outcomes above and beyond emotional intelligence or emotional skills (Groves & Feyerherm, 2011; Moon, 2010; Yoo et al., 2006), at least two studies have provided negative results (with the CQ; see Ward et al., 2009, Studies 2 and 3). Given that cross-cultural adaptation is an emotional process, it is not surprising that constructs assessed by tests of 3C overlap with constructs related to emotional skills. Future studies need to examine further the degree to which the constructs

underlying 3C tests are independent of emotion-related skills, and under what conditions they overlap or do not.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Portions of this report were prepared with the support of research Grant W91WAW-08-C-0024 from the Army Research Institute, and from research Grant FA9550-09-1-0281 from the Air Force Office of Scientific Research.

References

Ali, A., van der Zee, K. I., & Sanders, G. (2003). Determinants of intercultural adjustment among expatriate spouses. *International Journal of Intercultural Relations*, *27*, 563-580.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Anderson, P. H., Lawton, L., Rexeisen, R. J., & Hubbard, A. C. (2006). Short-term study abroad and intercultural sensitivity: A pilot study. *International Journal of Intercultural Relations*, *30*, 457-469.

Ang, S., Van Dyne, L., & Koh, C. (2006). Personality correlates of the four-factor model of cultural intelligence. *Group & Organization Management*, *31*, 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*, 335-371.

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

Arasaratnam, L. A., & Banerjee, S. C. (2011). Sensation seeking and intercultural communication competence: A model test. *International Journal of Intercultural Relations*, *35*, 226-233.

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

Atshuler, L., Sussman, N. M., & Kachur, E. (2003). Assessing changes in intercultural sensitivity among physician trainees using the Intercultural Development Inventory. *International Journal of Intercultural Relations*, *27*, 387-401.

Bakker, W., van der Zee, K. I., & van Oudenhoven, J. P. (2003). Individuele verschillen in reacties van migranten ten opzichte van culturele adaptatie [Individual differences in migrants' attitudes towards cultural adaptation]. *Nederlands Tijdschrift voor de Psychologie*, *58*, 81-94.

Bakker, W., van der Zee, K. I., & van Oudenhoven, J. P. (2006). Personality and Dutch emigrants' reactions to acculturation strategies. *Journal of Applied Social Psychology*, *36*, 2864-2891.

Bennett, M. J. (1986). Toward ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.), *Cross-cultural orientation: New conceptualizations and applications* (pp. 27-70). New York, NY: University Press of America.

Bhawuk, D. P. S. (1998). The role of culture theory in cross-cultural training: A multimethod study of culture-specific, culture-general, and culture theory-based assimilators. *Journal of Cross-Cultural Psychology*, *29*, 630-655.

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

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.

Bobowik, M., van Oudenhoven, J. P., Basabe, N., Telletxea, S., & Paez, D. (2011). What is the better predictor of student's personal values: Parent's values or student's personality? *International Journal of Intercultural Relations*, *35*, 488-498.

Brislin, R. W. (1993). *Understanding culture's influence on behavior*. Fort Worth, TX: Harcourt Brace Jovanovich.

Brislin, R. W., Cushner, K., Cherie, C., & Yong, M. (1986). *Intercultural interactions. A practical guide*. Beverly Hills, CA: Sage.

Cassiday, P. A. (2005). Expatriate leadership: An organizational resource for collaboration. *International Journal of Intercultural Relations*, *29*, 391-408.

Chen, A. S.-y., Lin, Y.-c., & Sawangpattanakul, A. (2011). The relationship between cultural intelligence and performance with the mediating effect of culture shock: A case from Philippine laborers in Taiwan. *International Journal of Intercultural Relations*, *35*, 246-258.

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

Chen, X.-P., Liu, D., & Portnoy, R. (2011). A multilevel investigation of motivational cultural intelligence, organizational diversity climate, and cultural scales: Evidence from U.S. real estate firms. *Journal of Applied Psychology*. Advance online publication. doi:10.1037/a00246

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

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

Dewaele, J.-M., & van Oudenhoven, J. P. (2009). The effect of multilingualism/multiculturalism on personality: No gain without pain for Third Culture Kids? *International Journal of Multilingualism*, *6*, 443-459.

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

Elenkov, D. S., & Manev, I. M. (2009). Senior expatriate leadership's effects on innovation and the role of cultural intelligence. *Journal of World Business*, *44*, 357-369.

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.

Fischer, R. (2011). Cross-cultural training effects on cultural essentialism beliefs and cultural intelligence. *International Journal of Intercultural Relations*, *35*, 767-775.

Gamst, G. C., Liang, C. T. H., & Der-Karabetian, A. (2011). *Handbook of multicultural measures*. Thousand Oaks, CA: Sage.

Goldstein, D. L., & Smith, D. H. (1999). The analysis of the effects of experiential training on sojourner's cross-cultural adaptability. *International Journal of Intercultural Relations*, *23*, 157-173.

Graf, A., & Harland, L. K. (2005). Expatriate selection: Evaluating the discriminant, convergent, and predictive validity of five measures of interpersonal and intercultural competence. *Journal of Leadership & Organizational Studies*, *11*(2), 46-62.

Greenholtz, J. F. (2005). Does intercultural sensitivity cross cultures? Validity issues in porting instruments across languages and cultures. *International Journal of Intercultural Relations*, *29*, 73-89.

Greenholtz, J. F., & Kim, J. (2009). The cultural hybridity of Lena: A multi-method case study of a third culture kid. *International Journal of Intercultural Relations*, *33*, 391-398.

Groves, K. S., & Feyerherm, A. E. (2011). Leader cultural intelligence in context: Testing the moderating effects of team cultural diversity on leader and team performance. *Group & Organization Management*, *36*, 535-566.

Gudykunst, W. B., Hammer, M. R., & Wiseman, R. (1977). An analysis of an integrated approach to cross-cultural training. *International Journal of Intercultural Relations*, *1*(2), 99-110.

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

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

Herfst, S. L., van Oudenhoven, J. P., & Timmerman, M. E. (2008). Intercultural effectiveness training in three Western immigrant countries: A cross-cultural evaluation of critical incidents. *International Journal of Intercultural Relations*, *32*, 67-80.

Hodges, N., Watchravesringkan, K., Karpova, E., Hegland, J., O'Neal, G., & Kadolph, S. (2011). Collaborative development of textile and apparel curriculum designed to foster student's global competence. *Family and Consumer Sciences Research Journal*, *39*, 325-338.

Houtz, J. C., Ponterotto, J. G., Burger, C., & Marino, C. (2010). Problem-solving style and multicultural personality dispositions: A study of construct validity. *Psychological Reports*, *106*, 927-938.

Imai, L., & Gelfand, M. J. (2010). The culturally intelligent negotiator: The impact of cultural intelligence (CQ) on negotiation sequences and outcomes. *Organizational Behavior and Human Decision Processes*, *112*, 83-98.

Jackson, D. N. (1974). *Personality research form manual*. New York, NY: Research Psychologists Press.

Kelley, C., & Meyers, J. E. (1987). *Cross-Cultural Adaptability Inventory manual*. Minneapolis, MN: National Computer Systems.

Kiuchi, A. (2006). Independent and interdependent self-construals: Ramifications for a multicultural society. *Japanese Psychological Research*, *48*, 1-16.

Klein, J. (1994). Effects of didactic teaching on Canadian cross-cultural sensitivity. *Psychological Reports*, *75*, 801-802.

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

Koester, J., & Olebe, M. (1988). The behavioral assessment scale for intercultural communication effectiveness. *International Journal of Intercultural Relations*, *12*, 233-246.

Korzilius, H., van Hooft, A., Planken, B., & Hendrix, C. (2011). Birds of different feathers? The relationship between multicultural personality dimensions and foreign language mastery in business professionals working in a Dutch agricultural multinational. *International Journal of Intercultural Relations*, *35*, 540-553.

LaFromboise, T., Coleman, H. L. K., & Gerton, J. (1993). Psychological impact of biculturalism: Evidence and theory. *Psychological Bulletin*, *114*, 395-412.

Lee, L.-Y., & Sukoco, B. M. (2010). The effects of cultural intelligence on expatriate performance: The moderating effects of international experience. *The International Journal of Human Resource Management*, *21*, 963-981.

Leone, L., Lucidi, F., Ercolani, A. P., & Presaghi, F. (2003). Versione italiana del Multicultural Personality Questionnaire (MPQ). *Bollettino de Psicologia Applicata*, *240*, 27-35.

Leone, L., van der Zee, K. I., van Oudenhoven, J. P., Perugini, M., & Ercolani, A. P. (2005). The cross-cultural generalizability and validity of the Multicultural Personality Questionnaire. *Personality and Individual Differences*, *38*, 1449-1462.

Leong, C.-H. (2007). Predictive validity of the Multicultural Personality Questionnaire: A longitudinal study on the sociopsychological adaptation of Asian undergraduates who took part in a study abroad program. *International Journal of Intercultural Relations*, *31*, 545-559.

Long, J.-h., Yan, W.-h., & van Oudenhoven, J. P. (2009). Cross-cultural adaptation of Chinese students in the Netherlands. *US-China Education Review*, *6*(9), 1-10.

Mahon, J. (2009). Conflict style of cultural understanding among teachers in the Western United States: Exploring relationships. *International Journal of Intercultural Relations*, *33*, 46-56.

Matsumoto, D., LeRoux, J. A., Bernhard, R., & Gray, H. (2004). Unraveling the psychological correlates of intercultural adjustment potential. *International Journal of Intercultural Relations*, *28*, 281-309.

Matsumoto, D., LeRoux, J. A., Iwamoto, M., Choi, J. W., Rogers, D., Tatani, H., & Uchida, H. (2003). The robustness of the Intercultural Adjustment Potential Scale (ICAPS). *International Journal of Intercultural Relations*, *27*, 543-562.

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.

Matsumoto, D., LeRoux, J. A., Robles, Y., & Campos, G. (2007). The Intercultural Adjustment Potential Scale (ICAPS) predicts adjustment above and beyond personality and general intelligence. *International Journal of Intercultural Relations*, *31*, 747-759.

Molinsky, A. L., Krabbenhoft, M. A., Ambady, N., & Choi, Y. S. (2005). Cracking the nonverbal code: Intercultural competence and gesture recognition across cultures. *Journal of Cross-Cultural Psychology*, *36*, 380-395.

Montagliani, A., & Giacalone, R. A. (1998). Impression management and cross-cultural adaptation. *Journal of Social Psychology*, *138*, 598-608.

Moon, T. (2010). Emotional intelligence correlates of the four-factor model of cultural intelligence. *Journal of Managerial Psychology*, *25*, 876-898.

Munroe, A., & Pearson, C. (2006). The Munroe Multicultural Attitude Scale Questionnaire: A new instrument for multicultural studies. *Educational and Psychological Measurement*, *66*, 819-834.

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*, 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*, 333-347.

Paige, R. M., Jacobs-Cassuto, M., & Yershova, Y. A. (2003). Assessing intercultural sensitivity: An empirical analysis of the Hammer and Bennett Intercultural Development Inventory. *International Journal of Intercultural Relations*, *27*, 467-486.

Paige, R. M., Jacobs-Cassuto, M., Yershova, Y., & DeJaeghere, J. G. (1999, April). *Assessing intercultural sensitivity: A validation study of the Hammer and Bennett (1998) Intercultural Development Inventory*. Paper presented at the International Academy of Intercultural Research, Kent State University, Kent, OH.

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, D. E. Wiley, & S. Messick (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67-88). Mahwah, NJ: Lawrence Erlbaum.

Pedersen, P. J. (2010). Assessing intercultural effectiveness outcomes in a year-long study abroad program. *International Journal of Intercultural Relations*, *34*, 70-80.

Ponterotto, J. G. (2010). Qualitative research in multicultural psychology: Philosophical underpinnings, popular approaches, and ethical considerations. *Cultural Diversity and Ethnic Minority Psychology*, *16*, 581-589.

Ponterotto, J. G., Costa-Wofford, C. I., Brobst, K. E., Spelliscy, D., Kacanski, J. M., Scheinholtz, J., & Martines, D. (2007). Multicultural personality dispositions and psychological well-being. *Journal of Social Psychology*, *147*, 119-135.

Ponterotto, J. G., Rieger, B. P., Barrett, A., & Sparks, R. (1994). Assessing multicultural counseling competence: A review of instrumentation. *Journal of Counseling & Development*, *72*, 316-322.

Ponterotto, J. G., & Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Perceptual and Motor Skills*, *105*, 997-1014.

Ponterotto, J. G., Ruckdeschel, D. E., Joseph, A. C., Tennenbaum, E. A., & Bruno, A. (2011). Multicultural personality dispositions and trait emotional intelligence: An exploratory study. *Journal of Social Psychology*, *151*, 556-576.

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*, *25*, 615-621.

Ramsey, J. R., Leonel, J. N., Gomes, G. Z., & Monteiro, P. R. R. (2011). Cultural intelligence's influence on international business traveler's stress. *Cross-Cultural Management: An International Journal*, *18*, 21-37.

Rockstuhl, T., Seiler, S., Ang, S., Van Dyne, L., & Annen, H. (2011). Beyond general intelligence (IQ) and emotional intelligence (EQ): The role of cultural intelligence (CQ) on cross-border leadership effectiveness in a globalized world. *Journal of Social Issues*, *67*, 825-840.

Ruben, B. D. (1976). Assessing communication competency for intercultural adaptation. *Group and Organization Studies*, *1*(3), 334-354.

Ruben, B. D., & Kealey, D. J. (1979). Behavioral assessment of communication competency and the prediction of cross-cultural adaptation. *International Journal of Intercultural Relations*, *3*, 15-47.

Rushton, J. P., & Irwing, P. (2009). A general factor of personality in the Comrey Personality Scales, the Minnesota Multiphasic Personality Inventory-2, and the Multicultural Personality Inventory. *Personality and Individual Differences*, *46*, 437-442.

Savicki, V., Downing-Burnette, R., Heller, L., Binder, F., & Suntinger, W. (2004). Contrasts, changes, and correlates in actual and potential intercultural adjustment. *International Journal of Intercultural Relations*, *28*, 311-329.

Scherer, M., Maddux, J., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. (1982). The Self-Efficacy Scale: Construction and validation. *Psychological Reports*, *51*, 663-671.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1-65). New York, NY: Academic Press.

Simkhovych, D. (2009). The relationship between intercultural effectiveness and perceived project team performance in the context of international development. *International Journal of Intercultural Relations*, *33*, 383-390.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*, 526-537.

Straffon, D. A. (2003). Assessing the intercultural sensitivity of high school students attending an international school. *International Journal of Intercultural Relations*, *27*, 487-501.

Suanet, I., & van de Vijver, F. J. R. (2009). Perceived cultural distance and acculturation among exchange students in Russia. *Journal of Community & Applied Social Psychology*, *19*, 182-197.

Templer, K. J., Tay, C., & Chandrasekar, N. A. (2006). Motivational cultural intelligence, realistic job preview, realistic living conditions preview, and cross-cultural adjustment. *Group & Organization Management*, *31*, 154-173.

Triandis, H. C., Brislin, R. W., & Hui, C. (1988). Cross-cultural training across the individualism-collectivism divide. *International Journal of Intercultural Relations*, *12*, 269-289.

Uhes, M. J., & Shybut, J. (1971). Personal orientation inventory as a predictor of success in peace corps training. *Journal of Applied Psychology*, *55*, 498-499.

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*, 283-303.

van der Zee, K. I., & van der Gang, I. (2007). Personality, threat, and affective responses to cultural diversity. *European Journal of Personality*, *21*, 453-470.

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*, 291-309.

van der Zee, K. I., & van Oudenhoven, J. P. (2001). The Multicultural Personality Questionnaire: Reliability and validity of self- and other ratings of multicultural effectiveness. *Journal of Research in Personality*, *35*, 278-288.

van der Zee, K. I., van Oudenhoven, J. P., & de Grijs, E. (2004). Personality, threat, and cognitive and emotional reactions to stressful intercultural situations. *Journal of Personality*, *72*, 1069-1096.

van der Zee, K. I., Zaal, J. N., & Piekstra, J. (2003). Validation of the Multicultural Personality Questionnaire in the context of personnel selection. *European Journal of Personality*, *17*, S77-S100.

van Oudenhoven, J. P., Mol, S., & van der Zee, K. I. (2003). Study of the adjustment of Western expatriates in Taiwan ROC with the Multicultural Personality Questionnaire. *Asian Journal of Social Psychology*, *6*, 159-170.

van Oudenhoven, J. P., Timmerman, M. E., & van der Zee, K. I. (2007). Cross-cultural equivalence and validity of the Multicultural Personality Questionnaire in an intercultural context. *Journal of Intercultural Communication*, *13*, 51-65.

van Oudenhoven, J. P., & van der Zee, K. I. (2002). Predicting multicultural effectiveness of international students: The Multicultural Personality Questionnaire. *International Journal of Intercultural Relations*, *26*, 679-694.

Wang, Y.-W., Davidson, 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*, 221-234.

Ward, C., Fischer, R., Lam, F. S. Z., & Hall, L. (2009). The convergent, discriminant, and incremental validity of scores on a self-report measure of cultural intelligence. *Educational and Psychological Measurement*, *69*, 85-105.

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

Weldon, D. E., Carlston, D. E., Rissman, A. K., Slobodin, L., & Triandis, H. C. (1975). A laboratory test of effects of culture assimilator training. *Journal of Personality and Social Psychology*, *32*, 300-310.

Yan, W.-h. (2009). The localization of Western personality instruments: In the cultural and methodological perspectives—As illustrated by the Multicultural Personality Questionnaire. *Psychological Science (China)*, *32*, 932-935.

Yoo, S. H., Matsumoto, D., & LeRoux, J. A. (2006). Emotion regulation, emotion recognition, and intercultural adjustment. *International Journal of Intercultural Relations*, *30*, 345-363.

Yuen, C. Y. M. (2010). Dimensions of diversity: Challenges to secondary school teachers with implications for intercultural teacher education. *Teaching and Teacher Education*, *26*, 732-741.