



## Intercultural Development Inventory (IDI): Independent Review

**Submitted By:**  
Andrew Wiley  
917.885.0858  
[Awiley@acsventures.com](mailto:Awiley@acsventures.com)

## Contents

Introduction .....	3
Validity .....	4
Content-Oriented Evidence .....	5
Background .....	5
Findings .....	5
Evidence regarding Response Process .....	6
Findings .....	6
Evidence regarding internal structure .....	6
Findings .....	7
Evidence regarding relationships with conceptually related constructs .....	7
Findings .....	7
Evidence Based on Consequences of Tests .....	7
Findings .....	8
Other research on the IDI .....	8
Bennett 2009 .....	8
Matsumoto 2013 .....	8
Use of IDI to Inform Selection .....	9
References .....	12

## Introduction

IDI, LLC (IDI) requested an independent review of the current state of the validity evidence associated with the use of *IDI Inventory* test scores. The IDI Inventory is designed to assess intercultural competence, defined as the “capability to shift cultural perspective and appropriately adapt behavior to cultural differences and commonalities.” The domain of intercultural competence has received attention recently. As more and more education and workplace settings include people from multiple cultures and backgrounds, the ability to work with people from different backgrounds becomes increasingly paramount. As will be discussed later in the report, intercultural competence has been identified as a critical capability in studies looking at students engaging in international travel or in semester abroad programs, as well as in job performance for managers who have diverse employees on their teams.

In response to this request, ACS Ventures, LLC (ACS) completed an independent review of the available validity evidence associated with the IDI Inventory and this report summarizes the results. ACS is a psychometric consulting company formed in 2016. ACS Ventures, LLC (ACS) was formed to address a need in the assessment community for design, operational support, and quality assurance. These needs are inclusive of assessment policy and practice in the education, credentialing, and workforce sectors. The ACS team is committed to applying its diverse experience in our work with organizations that is focused on the development of practical solutions that help ensure the reliability, validity, and fairness of our clients’ assessment programs.

ACS staff members have over 40 years of collective experience working with organizations in the education, workplace, and credentialing sectors. Their experience has included a review of comprehensive high-stakes statewide assessment programs, the independent evaluation of the validity and fairness of online assessment programs, and work setting standards in a wide variety of professional credentialing environments.

The review focused on evidence that was analyzed or collected throughout the test development, scoring, delivery, analyses and validation for the IDI Inventory. The project began with IDI staff members sharing any data, information or reports or other information that IDI considered to be critical for ACS to consider. ACS also conducted a literature review focused on the IDI Inventory in order to identify any external evaluations of the IDI Inventory that could be obtained as well identifying other researchers who have used the IDI Inventory in their work.

Throughout the process, ACS was guided primarily by the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Test Standards), and used the five defined sources of validity evidence as the expected validation framework. Specifically, evidence based on test content, response processes, internal structure, relations to other variables, and testing consequences was organized and reviewed relative to the intended purposes of scores from the instrument.

This report will walk through each of the five sources of evidence and review the data and artifacts identified for each. Before describing our findings, a few important limitations of this report are discussed here. First, this review is primarily focused on the psychometric validity of the IDI Inventory and was primarily based upon published research on the IDI Inventory. As such, this review did not include a review of IDI Inventory test content nor did it include a review of the soundness of the theory of action regarding intercultural sensitivity

(DMIS). In other words, our review provides information on how well the IDI Inventory adequately represents or is true to the model of intercultural sensitivity as defined by the developers. But it does not address the sensitivity of the model itself.

It should also be pointed out that this report is not designed to summarize every piece of research or work that has been completed with the IDI Inventory. Instead, the report summarizes the key findings of our review on whether or not the data and information uncovered support the use of IDI Inventory test scores for its currently defined purposes. The report also highlights areas of particular strength for the IDI Inventory, any areas where the evidence does not seem to be sufficient, conclusions reached by ACS, and a discussion of validation strategies to support expanded uses of scores. The review will also focus on some specific criticisms of the IDI Inventory that were observed in the literature.

## Validity

As was mentioned previously, the work conducted by ACS was guided by the Test Standards, and organized our review such that the data and information was considered as evidence for the five critical sources of validity evidence mentioned in the Test Standards. As outlined in the Test Standards, there are two critical concepts that should be highlighted before the report discusses what was found with the IDI Inventory research.

First, one critical characteristic is that validity should not be viewed as a dichotomous concept (i.e., the test either has validity or does not). Instead, validity is really a matter of degree, and as evidence supporting an intended interpretation or use of a test score accumulates, the degree of confidence in the validity of a given test score use can increase along with it.

A second important consideration is that validity is not something that belongs to a given test. Instead, validity must be considered within the given context and the intended use of a test score. The validity of any test does not get applied to that test, which then can be used regardless of how the test is being used. Instead, evidence must be accumulated for each use of a test and must help build an argument that the given use is appropriate. Certainly, as we reviewed the evidence related to the IDI, we saw how some pieces of evidence were appropriate and could be considered across multiple different uses. However, in other scenarios, as different uses for a test are introduced, it would be appropriate to consider what new evidence would be appropriate.

Kane (1992, 2006) is one of the leading researchers within the field of validity and has been one of the leading voices advocating the two critical characteristics described above. Kane discusses validity as building an argument to support the validity of the given test use. Rather than trying to identify a single study or a single piece of evidence, Kane urges test publishers to consider building a comprehensive argument to support the use of their test through multiple different studies. Each study should be designed to answer specific questions about the use of the test, and when viewing across all of the studies, a comprehensive picture can be drawn in support of the test.

As was mentioned in the Introduction, the IDI Inventory is designed as a measure of intercultural competence. The review conducted by ACS did not focus solely on one or two uses of the IDI Inventory. For example, IDI Inventory may be used in a program at a university designed to help support students who plan to travel

abroad for a semester. Other programs may use the IDI Inventory to help identify the needs for training within a team in an organization. The type of reports and information that would be needed to support the validity argument for each of these uses would need to be slightly different. This review was designed to evaluate how well the IDI Inventory captured the construct of intercultural competence; rather than a specific use. At the conclusion of this report, we will also review the evidence that would be appropriate to gather in the event that the IDI Inventory were to be used in other scenarios the extend the interpretation and use beyond the original intent, such as the selection of students or in the workplace to select employees.

## Content-Oriented Evidence

### Background

Within the Test Standards, the discussion of content-oriented evidence is based on the concept that important validity evidence is required that will examine the relationship between the construct being measured and the development of the test. Standard 1.11 states that “the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent” (pg. 26).

In addition to the standards for content-oriented evidence, the Test Standards also describe series of focused on the appropriate design and development of an assessment. These standards are designed to encourage a rigorous process in the development of both the specification and the items that appear on the test. The development of the specifications and the items should support the argument that the test is an appropriate representation of the construct being measured. For example, in Standard 4.8, the Test Standards recommends that experts judges be involved in the development of items and that they be included in the review of items as well as any scoring criteria. Another example can be observed in standard 4.12 which recommends that test developers “document the extent to which the content domain of a test represents the domain defined in the test specifications” (pg. 89).

### Findings

In our review of the IDI Inventory, the primary focus when looking at content-related evidence was to review the development process to ascertain how the IDI Inventory was developed, who was involved in the process, and what procedures were followed to develop the IDI. In a journal article by Hammer, Bennett, and Wiseman (2003), the process for creating the IDI Inventory, along with the theoretical underpinnings of the IDI Inventory are described in detail. The IDI Inventory is grounded within the Developmental Model of Intercultural Sensitivity (DMIS) initially developed by Bennett (1986; 1993). Within the article by Hammer et al. (2003), the model is fully described and the relationship to the IDI Inventory is fully documented. Hammer et al (2003) also described the process and people involved in the creation of the IDI Inventory.

As outlined in the article, the process for creating the IDI Inventory was consistent with the expectations described in the standards and are supportive of the overall validity of the IDI Inventory. As was already mentioned, this report is not designed to summarize data and information that is available elsewhere. Instead, the Findings section includes a summary of some of the key activities that were identified that support our conclusion for each source of validity evidence. Among the key activities observed were:

- The IDI Inventory was initially developed by gathering information and statements from a wide variety of speakers who represented different degrees of intercultural competence which helps supports the authenticity of the prompts used on the IDI Inventory.
- The IDI Inventory items followed a rigorous review and tryout process before they were eventually incorporated into the IDI Inventory. The reviews included independent experts who were knowledgeable in the concept of intercultural competence.
- All items were field tested and the statistical performance of the items and test as a whole were investigated using representative samples of test-takers.

## Evidence regarding Response Process

The Test Standards also recommends that data and information be collected that is focused on the response or cognitive process that test takers engage in when completing the assessment. Standard 1.12 states that if the rationale for score interpretations for a given use depends on the premise about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of these premises should be provided” (pg. 26). In many cases, evidence can be supplied here such as cognitive interviews or think aloud protocols with test takers who review test items and explain the process they followed when determining their answer. This process can also be reviewed through the independent review of test items or tasks by independent experts who help evaluate whether the intended purpose of each item is truly being tapped into by the item and if there are any construct irrelevant factors that could be impacting test taker performance.

## Findings

During our review of the IDI Inventory, our focus was on the activities associated with gaining a better understanding how test takers were responding to IDI test items. At this time, the IDI Inventory test items have not been reviewed through activities such as cognitive labs, interviews, or reviews typically conducted in alignment studies. The IDI Inventory test items were developed using a rigorous process, including a review of the items by individuals knowledgeable about intercultural competence who indicated whether the items were appropriate and representative of key features within the domain. While this review does not replace the need for cognitive labs or other similar activities, it does provide some evidence to support the cognitive processing test-takers would use while completing the IDI Inventory.

## Evidence regarding internal structure

When considering evidence based upon internal structure, the Test Standards emphasize the need for evidence that demonstrates that the test items and scores provided to test-takers are consistent with the strand structure of the test or the reporting structure of the test. For example, in an educational test with one total score, and four subscores, evidence would be expected to show that the items assigned to each of the four subscores was consistent with how the items performed. In Standard 1.13, the Test Standards indicate that “the relationship among test items or among parts of the test, evidence concerning the internal structure of the test should be provided” (pg. 27). For many testing programs, the evidence within this domain tends to investigate the factor structure of the assessment and investigating whether the factor structure is consistent with the reporting structure of the test. This evidence can also include data such as the correlation across subscores provided to test takers.

## Findings

The IDI Inventory has undergone multiple investigations focused on the internal factor structure. For example, Hammer (2011) conducted a confirmatory factor analysis (CFS) investigation into the IDI Inventory using more than 4,700 IDI Test takers. In general, the results from this study were consistent with the reporting structure used for the IDI Inventory.

A second area that was reviewed was the scoring procedures to evaluate how the IDI Inventory has used concepts such as “trailing orientations” or “developmental orientations.” This part of the IDI Inventory reporting process can provide valuable information to test-takers and help guide the interpretation of the IDI Inventory test scores. However, in many of these scenarios, the cut scores or scoring rules that are used by the IDI Inventory have not been developed through a rigorous process that helps ensure that the scoring rules are consistent with the requirements and needs of the IDI Inventory test users.

## Evidence regarding relationships with conceptually related constructs

When considering this area of validity evidence, most of the research is focused on the correlation of test scores with other similar assessments or measures. Test Standard 1.16 states that “evidence concerning the constructs represented by other variables, as well as their technical properties, should be presented or cited” (pg. 27). For example, research studies can be conducted that compare test scores to other similar test scores or measures that are similar. For example, educational tests may be compared to performance in school, such as grades in classes. Research in this area can also include research demonstrating that test scores are *not* related to other critical variables, such as social desirability.

## Findings

There have been numerous studies that investigated the relationship of IDI Inventory test scores with other similar constructs. For example, Paige, Jacobs-Cassuto, Yershova & DeJaeghere (2003) investigated the relationship of IDI Inventory test scores and found that test scores were correlated with other variables, such as the number of friends from other cultures, and time spent abroad. Hammer (2005) also found that IDI Inventory test scores were related to less intercultural anxiety with students as well as satisfaction with study abroad programs. Kruse, Didion, & Perzynski (2014) utilized the IDI Inventory in work they completed with health care workers and found that the IDI Inventory could be used as a measure of cultural competence with teachers and students. Overall, the data and evidence provided in this section is consistent with the expected industry standards.

## Evidence Based on Consequences of Tests

Test Standard 1.25 recommends that “an attempt should be made to investigate whether such consequences arise from the test’s sensitivity to characteristics other than those it is intended to assess” (pg. 30). For many programs, this research can include focus groups and interviews with test users to determine how the end users actually use the assessment and how the test scores are implemented. In other scenarios, an attempt is made to identify misuse of the test scores and to determine if the misuse of the test scores has caused harm to test-takers.

## Findings

For the IDI, the organization has provided numerous case studies that are focused on the appropriate implementation of the IDI Inventory and how the test scores can be used in organizations. These studies provide useful information to IDI test users and are consistent with the industry practices in this area.

## Other research on the IDI

In addition to the studies identified in earlier sections of this report, our review also included other studies that either reviewed the IDI or conducted research using the IDI Inventory. Some of these studies will be highlighted here along with a brief description of their data and conclusions.

### Bennett 2009

As was mentioned previously, the IDI Inventory is thoroughly grounded in Bennett's Developmental Model of Intercultural Sensitivity (Bennett 1986; 1993). Because of that, our review included a review of many reports written by Bennett, including a review in 2009 focused on the defining and measuring of intercultural learning. In the report, Bennett stated that the IDI Inventory sacrificed ideographic data over nomothetic data. In effect, Bennett is stating that because the IDI Inventory is used to categorize test takers into groups, it loses some information at the individual level. For example, if two individuals are both classified into the Minimization category, they are not necessarily the same, and the IDI is not able to track or describe these differences. It does not appear that Bennett's criticism of the IDI Inventory would be limited to solely the IDI Inventory, but would apply to any assessment that provides information on test takers and classifies them into groups. On the other hand, collecting individualized data for all test takers and having comprehensive information for each does not seem reasonable either, and making judgments about the individual or groups with that level of data might not be feasible.

The other statement that Bennett makes is that because it sacrifices ideographic information, the IDI Inventory subsequently overestimates the degree of candidates in the Minimization position. At this point, it is not clear why Bennett states that the minimization condition is overestimated with the IDI Inventory, and the article does not provide further details or data to support the statement. At this point, it does not appear that Bennett has provided further information or data to support his position, so it probably should be viewed as an unsupported assertion at this time.

### Matsumoto 2013

In a review article published in 2013, Matsumoto completed a review of 10 measures of intercultural competence. Within the review, Matsumoto identified ecological validity as the primary criterion used to evaluate the appropriateness of each test. For Matsumoto, ecological validity is a comprehensive term that primarily includes the relationship of the test scores with other measures, such as other external measures of intercultural competence. These studies could also include studies that show changes in test scores after training in intercultural competence and measures of intercultural competences across difference groups. The primary concern that was raised during our review of the Matsumoto article is that the definition of validity does not appear to be consistent with current industry standards. As was outlined in this report, the Test Standards, support the review of evidence across five key areas and evidence is expected within all five

areas. While certainly the evidence that Matsumoto labeled ecological validity are an important part of any review, it is not sufficient for a comprehensive review of the validity evidence of any program.

## Use of IDI to Inform Selection

An additional question that was explored during our review was the extent to which it would be appropriate to use the IDI Inventory as part of a selection process, whether it be for the selection of employees or for the selection of students into an educational program. As has been described in the report, the development procedures and the research surrounding the IDI Inventory provides a strong argument for the IDI Inventory as a valid measure of intercultural competence. However, as was also mentioned, validity is not something that gets assigned to a test, but is rather something that is evaluated as appropriate for test scores being used in a particular fashion. For that reason, it would be appropriate to consider three key questions when determining if the IDI would be appropriate for selection.

1. Can the construct of intercultural competence be considered an important knowledge, skill, or ability for an individual to succeed in a given position or succeed within a given educational program or job role?
2. Is the IDI Inventory a valid measure of intercultural competence?
3. Is there a demonstrated, predictive relationship between performance on the IDI Inventory and performance as an employee or student on criterion variables of successful performance in these settings?

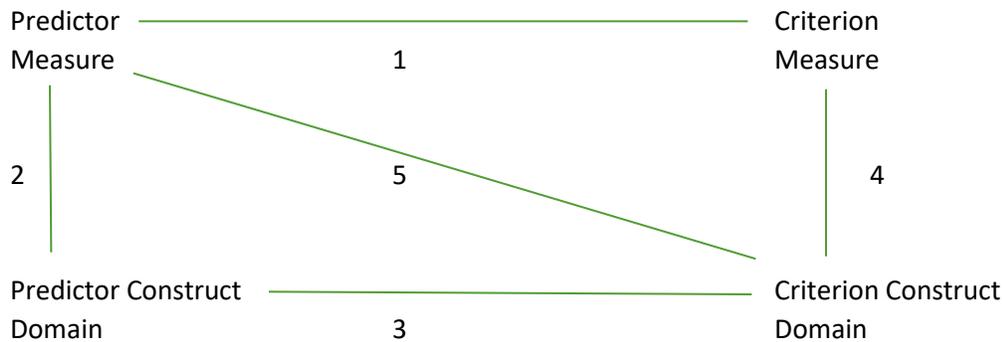
Based upon the review that ACS has completed, it does appear that there is strong support for #2, which is obviously a critical consideration. However, before adopting the IDI Inventory for use as a potential tool for informing selection decisions, the other two questions must also be considered. Because the use of any assessment within a selection environment significantly raises the stakes of that assessment and the potential legal risk, it is reasonable to assume that the validity evidence for the assessment will receive an increased degree of scrutiny.

As described in the Test Standards, the critical inference that must be supported is that “scores on the test can be used to predict subsequent job performance” (pg. 172). As illustrated in Figure 1 below, The Test Standards discuss five different types of linkages that can be considered when determining whether or not an assessment is appropriate to use in a selection environment. Within the Test Standards, the initial focus is on the two domains that are of interest, the *predictor construct domain* and the *criterion construct domain*. In the scenario where the IDI Inventory was being considered for use in the selection of employees within an organization, the construct of intercultural competence is the predictor construct domain and performance within a given position is the criterion construct domain.

In traditional selection research and validation, linkages 2 and 4 are similar to one another. In both cases, the critical question that is being asked is whether the particular test score or rating is a valid indicator of either the predictor construct domain or the criterion construct domain. In our scenario, the linkage illustrated in linkage 2 is the link between IDI Inventory scores and the construct of intercultural competence. The question being asked when considering linkage #2 is very similar to the question being asked in our #2 question above. Because of that, we believe there is strong evidence to support the assertion that the IDI Inventory is a reliable, valid, and fair measure of intercultural competence. Linkage #4 highlights the fact that there must be a relationship between the construct in question and the measure of job performance. In most scenarios, it

would be the responsibility of the employer or organization to provide some evidence to support the assertion that the criterion measures or job performance measures are an appropriate measure of job performance.

**Figure 1: Predictor and criterion relationship**



When considering how IDI Inventory scores can potentially be used to inform the selection process, it is important for us to highlight the linkages represented as linkage #1 and linkage #3. The Test Standards states that linkage #3 is usually identified through “theoretical or logical analysis” (pg. 172). Through a systematic review of the job in question, a construct such as intercultural competence can be identified as an essential skills or ability in this particular job. Once that is established, the linkage in #3 begins to form and the relationship between the predictor construct domain and the criterion construct domain begins to form. In the scenario for IDI Inventory test scores, linkage #3 can be satisfied through a systematic review that identifies intercultural competence as critical in a given job. A systematic review is usually completed through procedures such as a job or practice analysis that will include the collection of data on the knowledge, skills, and abilities required to perform in a given position as well as the judgment of experts who evaluate the data and determine the most appropriate weighting of the knowledge, skills, and abilities.

When data and evidence is available to support linkages #2, #3, and #4, a strong theoretical argument can be built to support the use of a predictor measure when selecting candidates for a given position. However, even with that information, it is also important that research be constructed that focuses on linkage #1 as well. The relationship represented by linkage 1 is the relationship investigated via a study to determine if there is a correlation between the criterion measure (i.e., job performance) and the predictor measure (i.e. the IDI Inventory).

Within this review, a number of studies have been reviewed that provided strong evidence to support linkage #2, the link between the construct of intercultural competence and the IDI Inventory. Other studies have also been observed that have supported linkage #1. Hammer (2011) conducted a study that investigated the relationship between the IDI Inventory test scores and performance as a hiring manager within a group or organizations. The study found that individuals at higher points on the IDI scale were more likely to recruit and select a diverse set of employees. The IDI has also demonstrated a correlation between IDI test scores and students’ behavior in study abroad programs. For example, students who scored higher on the IDI Inventory were more likely to have greater knowledge of the host culture in the program, and also were more likely to have a greater number of intercultural friends.

Because of the collection of studies, it does appear to be a reasonable assumption to think that the IDI Inventory could be a useful tool in a selection environment. However, even though it is a reasonable assumption, it would still be advisable for all parties involved in such a decision to carefully evaluate all of the evidence and determine how consistent the evidence is with their given scenario. While a relationship between IDI Inventory test scores and measures of job performance in some scenarios could be a reasonable assumption, it would still be advisable to complete the research associated with linkage #1, in particular at the beginning of using the IDI Inventory.

One of the challenges that exists with the use of any test score is how often the test score relationship needs to be investigated. As has been mentioned previously in this report, validity is not something assigned to a test, but is instead related to a test score within a given test use. Taken literally, this would imply that every unique usage of the IDI Inventory would require a subsequent study to demonstrate the linkage between IDI Inventory test scores and job performance. However, in many scenarios, when the knowledge and skills are reasonably similar to studies that have been completed in the past, this research can provide a sense of confidence that IDI Inventory test scores would be appropriate. This concept is characterized as validity generalization (see Schmidt & Hunter, 1977). Although there is not a clear delimiter or marker to identify when a new study would be required, some guidance regarding the need for new or additional studies can be found. For example, when new uses of IDI Inventory test scores are considered, a review should be completed that considers things like the similarity of job tasks and the environment for the job or educational setting, and a reasonable expert judgment, must be made that determines whether a new study is required. In scenarios that are closely parallel to previous research studies, test users can be reasonably confident that the use of IDI Inventory test scores is appropriate and supported. In other scenarios, when the specific job responsibilities are different, or the environment in which the tasks are being completed is different, it is advisable to move more cautiously, and to introduce the use of the IDI Inventory in a lower stakes manner and study the relationship within the current group of employees.

Another factor that should be considered is whether the IDI Inventory is used as the sole predictor of job performance or is being used within a battery of other assessments. When any assessment is being used within an overall battery of assessments, it becomes more difficult to tease out and investigate the impact and relationship of a single assessment. Nonetheless, it is still critical that evidence be available to support the comprehensive selection protocols, including each specific instrument. It is also important to consider how each assessment is being used in the given selection scenario. In the event that the test scores for the IDI Inventory are being used solely to provide some additional information about a candidate, but is not being used to select or eliminate candidates, the evidence will not need to be as extensive. However, when an assessment like the IDI Inventory is being used to select candidates (e.g. a minimum cut score is set), evidence to support this decision rule should be collected.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bennett, M.J. (1986). Towards ethnorelativism: A developmental approach to training for intercultural sensitivity. *International Journal of Intercultural Relations*, 10 (2), 179-196.
- Bennett, M.J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R.M. Paige (Ed.), *Education for the intercultural experience* (2nd ed.) (21-71). Yarmouth, ME: Intercultural Press.
- Bennett, M.J. (2009). Defining, measuring, and facilitating intercultural learning: a conceptual introduction to the *Intercultural Education* double supplement. *Intercultural Education*, Vol. 20, No. S1–2, S1–13.
- Hammer, M.R., Bennett, M.J. & Wiseman, R. (2003). The Intercultural Development Inventory: A measure of intercultural sensitivity. In R.M. Paige (Guest Editor), Special issue on the Intercultural Development Inventory. *International Journal of Intercultural Relations*, 27, 421-443.
- Hammer, M. R. (2005). *Assessment of the impact of the AFS study abroad experience: Executive Summary*. New York: AFS, International.
- Hammer, M.R. (2011). Additional cross-cultural validity testing of the Intercultural Development Inventory. *International Journal of Intercultural Relations*, 35, 474-487.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–535.
- Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport, CT: American Council on Education and Praeger.
- Kruse, J.A., Didio, J., Perzynski, K. (2014). Utilizing the Intercultural Development Inventory to develop intercultural competence. In SpringerPlus 3:34 available at <http://www.springerplus.com/content/3/1/334>
- Matsumoto, D. & Hwang, H.C. (2013). Assessing Cross-Cultural Competence: A Review of Available Tests. *Journal of Cross-Cultural Psychology*, 44(6), 849-873.
- Paige, R.M., Jacobs-Cassuto, M, Yershova, Y.A., & DeJaeghere, J. (2003). Assessing intercultural sensitivity: An empirical analysis of the Hammer and Bennett Intercultural Development Inventory. In R.M. Paige (Guest Editor). Special issue on the Intercultural Development Inventory. *International Journal of Intercultural Relations*, 27, 467-486.
- Schmidt, F. L. & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of applied psychology*, 62, 529-540.
- Sireci, S.G. (2013). Agreeing on Validity Arguments. *Journal of educational measurement*, 50, 99-104.

