Theory and intervention equivalence and bias: New constructs and adaptation of the systematic test of equivalence procedure

Lawrence H. Gerstein[1,2] (ID)

Abstract

The importance of cross-cultural validity, equivalence, and bias when engaged in cross-cultural and cultural activities (e.g., research, practice, training, supervision, consultation, teaching) are discussed. Two new forms of equivalence (Theory; Intervention) and bias (Theory; Intervention) also are introduced, as is an adaptation of the Systematic Test of Equivalence Procedure (STEP; Gerstein, L.H. 2018. Systematic test of equivalence procedure: New method to investigate cross-cultural validity. *Revista de Cercetare si Interventie Sociala, 62,* 278–293), which was originally created to enhance equivalence and reduce bias when conducting cross-cultural or cultural studies. Assuming future research supports the validity and relevance of the modified STEP approach, this strategy could offer researchers, clinicians, and others a tool to examine theory and intervention equivalence and bias in their professional activities. Recommendations for conducting projects to investigate the validity of employing STEP when examining the equivalence or bias of existing and future theories and intervention strategies are presented as well.

Keywords Cross-cultural validity · Equivalence · Bias · Systematic test of equivalence procedure

Cross-cultural validity is an infrequently discussed concept in the psychology literature. It is even less often tested and established in studies warranting its assessment. The consequences of this omission are far reaching. That is, our theories, scientific philosophy, research paradigms and procedures, measurement devices, applications, and training, education, and supervision paradigms, for example, are weakened and potentially jeopardized.

Few scholars have clearly defined cross-cultural validity, and no theories or models exist currently to explain this concept though there is a need for such schemata. Some researchers (Gerstein, 2011; Ægisdóttir, Gerstein, & Çinarbas, 2008) have defined this construct as establishing the relevance of concepts, methods, measures, and statistics when conducting research, and/or employing suitable counseling strategies or other interventions in various cultures. Moreover, it has been posited that cross-cultural validity is established when a construct operationalized in one culture (country), and the methodology and analyses utilized to evaluate this construct, are valid in a different culture (country) (Gerstein, Heppner, Ægisdóttir, Leung, & Norsworthy, 2009). There are many concepts linked to cross-cultural validity including equivalence and bias.

This article discusses the purpose of various forms of equivalence and bias, and the importance of these constructs to cross-cultural validity and research. It also introduces two new forms of equivalence (Theory and Intervention) and bias (Theory and Intervention). Further, it describes an adaptation of the Systematic Test of Equivalence Procedure (STEP; Gerstein, 2018), a method designed to enhance equivalence and reduce bias when performing cross-cultural activities. Lastly, as research studies have yet to be conducted on the two new types of equivalence and bias and also the validity of the adaptation of STEP introduced in this article, some recommendations are discussed to investigate these topics. Additionally, suggestions are offered on how to employ the STEP in the practice of psychology.

Types of Equivalence and Bias

Equivalence As others have claimed (van de Vijver, 2001; Ægisdóttir et al., 2008), equivalence is a significant concept in cross-cultural psychology to address. There are various kinds of equivalence, and depending on the target objectives,

✉ Lawrence H. Gerstein
  lgerstein@bsu.edu

1   College of Health, Ball State University, Room 449, Muncie, IN, USA

2   Department of Counseling Psychology, Social Psychology, and Counseling, Muncie, IN 47306, USA
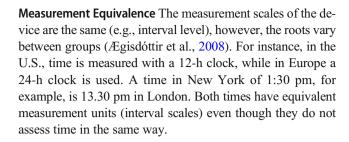
one or more must be present to establish evidence of cross-cultural validity (Gerstein, 2018). Definitions of different types of equivalence follow with two new kinds (Theory; Intervention) introduced in this article.

Construct Equivalence The target construct has a similar basic meaning in the cultures (countries) of interest (He & van de Vijver, 2012; Ægisdóttir et al., 2008). For instance, in a study by Bellare and Gerstein (2018), everyday stranger harassment (Bowman, 1993) was viewed similarly in India and the U.S. thereby demonstrating the construct equivalence of this concept in these two countries.

Structural equivalence is a specific form of construct equivalence (van de Vijver, 2001). It refers to when the construct is structurally comparable for the targeted culture (country) and when there is an equivalent meaning for the construct (Ægisdóttir et al., 2008). For instance, when studying the use of nonviolent behavior to resolve conflict among Hong Kong youth, a scale created in the U.S., the Teenage Nonviolence Test (TNT; Gerstein, Mayton, Hutchison, & Kirkpatrick, 2014; Mayton, Weedman, Sonnen, Grubb, & Hirose, 1999), was slightly modified for this sample, administered (Gerstein et al., 2018), and responses were subjected to principal components and confirmatory factor analyses. In general, the results supported the structural equivalence between the Hong Kong and U.S. versions of the TNT.

Method Equivalence The same approach to collect data (e.g., self-report, interviews) or method (e.g., field, laboratory) can be used to study the target construct in the culture (country) of interest (Gerstein, 2018). For example, Kim and Gerstein (2011) found a self-report scale, the Calling and Vocation Questionnaire (CVQ; Dik, Eldridge, Steger, & Duffy, 2012), could be given to Korean and U.S. college students. Thus, method equivalence was demonstrated. It is critical to establish this form of equivalence prior to doing a cross-cultural study. Some evidence exists indicating that using diverse methods (e.g., computer-assisted vs. paper scales) led to different results in cross-cultural studies (van de Vijver & He, 2016) though there is scant research or discussion of this finding in the literature (Gerstein, 2018).

If a scale is appropriate for use in two or more cultures (countries), it is also essential to establish that a similar item format can be utilized to demonstrate method equivalence (Gerstein, 2018). For example, Kim and Gerstein (2011) found they could employ a Likert format when administering the CVQ to Korean and U.S. students. It is critical to not assume individuals from different cultures (countries) can understand or relate to the same item response formats because people around the world are not all familiar with the same formats (He & van de Vijver, 2012; International Test Commission Guidelines for Translating and Adapting Tests, 2015).

Measurement Equivalence The measurement scales of the device are the same (e.g., interval level), however, the roots vary between groups (Ægisdóttir et al., 2008). For instance, in the U.S., time is measured with a 12-h clock, while in Europe a 24-h clock is used. A time in New York of 1:30 pm, for example, is 13.30 pm in London. Both times have equivalent measurement units (interval scales) even though they do not assess time in the same way.

Scalar [Metric] Equivalence Measures that are identical at the scalar level assess a construct with an equivalent ratio or interval scale for the targeted cultures (countries) and the origins of the measures are the identical (Ægisdóttir et al., 2008).

Linguistic Equivalence This kind of equivalence involves the item wording, form, meaning, and structure when various language versions of a scale exist (Ægisdóttir et al., 2008). It also involves the items' reading difficulty (Lonner, 1985). Linguistic equivalence is shown when the language of each scale is similar given the criteria just discussed (Gerstein, 2011).

Theory (Model or Paradigm) Equivalence Theory equivalence is a new form of equivalence introduced in this article. Previously, no such concept existed limiting psychologists systematically understanding and investigating a theory's cross-culturally validity. Theory equivalence is now defined as follows: A targeted theory, model, or paradigm has a comparable functionality, structure, and meaning to guide the activity (e.g., research, practice, training, supervision, consultation, teaching) in the culture (country) of interest. In specific, to demonstrate this type of equivalence, the components of the theory are relevant and applicable to the particular phenomena, population, context, and/or situation in the targeted culture (country) including the (a) philosophical assumptions, (b) constructs, (c) logic, (e) operational definitions, (f) hypotheses, (g) principles or links between the structural association between two or more concepts or constructs, (h) observations, and (i) explanatory power.

While this form of equivalence shares some similarities with cultural validity and the Cultural Lens Approach (CLA; Hardin, Robitschek, Flores, Navarro, & Ashton, 2014), it also is distinct. Leong and Brown's (1995) definition of cultural validity is quite broad, focusing on the presence of "construct, concurrent, and predictive validity of theories and models across cultures" (p. 144). Later, Quintana, Troyano, and Taylor (2001) expanded this definition to, "how constructs are operationalized, participants are recruited, hypotheses are formulated, study procedures are adapted, responses are analyzed, and results are interpreted for a particular cultural group" (p. 617) along with the utility of the research (e.g., instructional, practical). Historically, however, most scholars have only examined the cultural validity of scales (Hardin et al., 2014) and the theories underlying these devices (e.g., Fouad & Kantamneni, 2010).

Like the concept, Theory Equivalence, cultural validity and, in general, the CLA incorporate the importance of determining if constructs, operational definitions, and hypotheses linked to an existing framework are appropriate for a specific purpose. Theory Equivalence, however, focuses more closely on a larger number (e.g., philosophical assumptions, logic, structural connection between two or more constructs) and more in-depth assessment of the characteristics and components of a framework's infrastructure.

A poignant illustration of scholars pursuing this type of equivalence though they did not label it as such is the evolution and adaptation of the Contact Hypothesis (Allport, 1954) later known as the Intergroup Contact Theory (Pettigrew, 1998). This hypothesis stipulates four components essential for optimizing interpersonal contact between out-group and in-group members: equal-status between the different groups, common goals of the groups, cooperation between the groups, and support for intergroup contact by authorities, law, or norms. Based on many studies designed to test this hypothesis, Pettigrew (1998) introduced the Intergroup Contact Theory. Drawing on research involving individuals' intergroup contact within European countries and between people of different European countries, he posited, in part, there were four interconnected processes that operate through intergroup contact and mediate individuals' attitudes among other outcomes toward the out-group. These processes included learning about the outgroup, changing one's behavior, generating affective ties (e.g., increased empathy) between in- and out-group members, and ingroup reappraisal (members modifying their perceptions of their ingroup that could result in less provincial views of the out-group).

The research since 1954 on intergroup contact between, for example, persons in different cultures and countries, has led to adapting the Intergroup Contact Theory in terms of some its philosophical assumptions, constructs, logic, hypotheses, structural links among the constructs, and explanatory power. This outcome has increased the theory's relevance when explaining the nature and outcomes of interactions between individuals of diverse groups. Stated another way, this outcome has contributed to establishing evidence for the theory equivalence of the Intergroup Contact Theory for different groups of people and in diverse contexts and situations.

Intervention (Strategy) Equivalence Intervention equivalence is another new type of equivalence introduced in this article. This form of equivalence has not been discussed in the literature hindering psychologists ability to systematically comprehend and examine an intervention's cross-cultural validity. In this article, intervention equivalence is defined as an intervention(s) created in one culture (country) is applicable to the target culture (country). The intervention(s) also can be applied in that culture (country) with specific populations, situations, and/or contexts in a similar way. These interventions

might be used, for instance, in counseling, consultation, supervision, education, training, prevention, social justice, program development and implementation, and conflict prevention and resolution work. In particular, the following aspects of the intervention(s) must be comparable to establish this kind of equivalence: (a) philosophical assumptions, (b) logic, (c) operational definitions, (e) content, (d) target population, context, and/or situation where applied, (e) rationale for implementation, (f) when it is employed, (g) how it is delivered, (h) parameters of utilization, (i) expected outcomes, and (j) ethical boundaries and constraints.

Some examples of individuals adapting interventions first created in the U.S. to be more culturally valid to other samples appear in the Cognitive Behavioral Therapy (CBT) literature. While these people did not discuss equivalence in their rationale for modifying CBT strategies, it is rather apparent they addressed establishing what I have termed, intervention equivalence. For example, Naeem et al. (2015), Pakistani scholars, adapted CBT interventions to be applicable when working with South Asian Muslim patients. This was done to counter what they and others (Hays, 2009) claimed to be an individualistic perspective of CBT, and as such to offer Pakistani patients from a collectivist culture, interventions that were more effective. Similar to CBT in the U.S. and based on their research and clinical experience, Naeem et al. (2015) recommended that when confronting dysfunctional beliefs and cognitive errors, psychologists should keep in mind Pakistanis have a desire to please people close to them, succumb to the requests of people they love, and meet the needs of their family at the expense of their own needs. To determine if such beliefs are functional or not, Naeem et al. (2015) recommended determining how a client's family or community perceived these needs in terms of their functionality. Naeem et al. (2015) also suggested how to modify strategies to help Pakistanis distinguish between thoughts and emotions, and make homework assignments more culturally relevant.

There are many other examples of CBT adapted so intervention equivalence could be established, including studies in Hong Kong (e.g., Lau, Chan, Li, & Au, 2010; Wong, 2008). The strategies used in these studies were applied in a similar way to the U.S. with comparable clients. The rationale for these interventions was consistent with use in the U.S. The content of the techniques was slightly changed though congruent with basic assumptions and strategies of CBT.

Bias Another concept related to exploring and establishing cross-cultural validity is bias. Bias is inversely correlated with equivalence (Ægisdóttir et al., 2008). As equivalence increases, bias decreases, and vice versa. Bias, therefore, is a nuisance factor restricting the similarity of observations between cultural groups (Ægisdóttir et al., 2008). Three types of bias reported in the literature and two new forms (Theory; Intervention) introduced in this article are defined below.

Construct Bias This bias occurs if a construct assessed in its entirety is not equivalent between cultures (countries). For example, people from India and Brazil might not view anxiety similarly. If a study ignored this distinction, construct bias would be a concern. Factors contributing to this bias are not thoroughly capturing how a construct is defined for different cultures (countries), inconsistent investigation of the construct between cultures (countries), and incompatible item content on multilingual versions of a scale (cf. Ægisdóttir et al., 2008).

Method Bias This bias is linked to features of a scale and/or how a scale is administered (van de Vijver, 2001; Ægisdóttir et al., 2008). One example of this bias is biased scores. Another example is if a self-report scale assessing altruism was not appropriate to give to Iraqis because they preferred being interviewed about this topic. Proceeding with giving the scale, therefore, would introduce method bias. It is also critical to consider the item response format before using it in various cultures (countries) (Gerstein, 2018). If this is not done, method bias can occur. For instance, if a forced choice format (e.g., yes or no) was not the right one to employ with Laotians because they preferred a Likert format, using a forced choice format would introduce in the study instrument bias, a form of method bias. There are other kinds of method bias that can occur in cross-cultural and cultural studies that result from participants' ways of responding to response formats, including, for example, a midpoint (e.g., preference for neutral) response style for Likert items, an acquiescence response style (e.g., inclination to agree), an extreme response style (e.g., strongly disagree, strongly agree), and a modesty effect kind of response.

Item Bias This bias happens when an item(s) is understood differently in the culture (country) being studied (He & van de Vijver, 2012). Thus, this bias is connected to measurement of the item itself (Ægisdóttir et al., 2008). Formulating an item poorly, including items where the content might not be appropriate or relevant for the cultures (countries) being studied, and not translating it correctly can produce item bias (e.g., Malpass & Poortinga, 1986).

Theory (Model; Paradigm) Bias This also is a new construct introduced in this article, and not previously examined or understood in a systematic way. It is defined as a theory created in one culture (country) is not applicable to another culture (country). This bias happens if the functionality, structure, and meaning of a theory guiding an activity (e.g., research, practice, training, supervision, teaching, consultation) is inconsistent with the target culture (country). This bias occurs if one or more of the following components of a theory are incompatible with a specific phenomenon, population, context, and/or situation in the culture (country) of interest: (a) philosophical assumptions, (b) constructs, (c) logic, (e) operational definitions, (f) hypotheses, (g) principles or links between the structural association between two or more concepts or constructs, (h) observations, and (i) explanatory power. The extent of this bias varies by how many and to what degree the components are incompatible between the cultures (countries).

For some time, psychologists have studied and integrated into their practice the construct, well-being. One model of well-being created in the U.S., Ryff's (1989) Subjective Psychological Well-Being model, has garnered a great deal of attention worldwide. It will be assessed for a theory bias. Ryff's model integrates assumptions, for example, from humanistic, existential, and developmental paradigms. Further, the foundation is grounded in what some scholars (e.g., Norsworthy, 2017) call the minority world or Global North. The six dimensions of Ryff's (1989) model reflect these ideologies: self-acceptance, positive relations with others, autonomy (independence), environmental mastery, purpose in life, and personal growth.

Many studies have tested Ryff's model relying on his Scales of Psychological Well-Being (SPWB, Ryff, 1989). While these studies in many instances provided support for Ryff's six dimensions, there are also highly inconsistent results with respect to the model's factor structure, especially in research done outside the U.S. Explanations for this inconsistency include characteristics of the samples, the usage of different versions of the SPWB (e.g., short versus longer forms), and the rigor of the translation and back-translation processes if used at all.

Although these are viable reasons, questions remain about the conceptualization and structure of Ryff's (1989) model, particularly the equivalence of the dimensions in different countries as these varied somewhat. Further, as others have said (Choi & Choi, 2016; Joshanloo, 2014), it is conceivable differences in the conceptualization and presence of the six dimensions may vary by individuals' religious and spiritual beliefs, especially when comparing how well-being is understood in Western, Eastern, and Middle Eastern cultures and countries. In the latter two cultures and countries, Hinduism, Buddhism, Taoism, Confucianism, Sufism, and Islam are ancient religious and spiritual belief systems intricately intertwined with daily life. There are major distinctions between Ryff's (1989) model and how these systems conceptualize well-being. In these systems, in general, well-being is experienced by "letting go," not controlling or mastering one's situation, restricting a focus on, and transcending the self, pursuing individual, group, and societal harmony, and embracing spiritual and religious beliefs and actions. Ryff's model, in contrast, emphasizes self-determination, actively striving to change oneself, and mastery of one's experience, relationships, and the environment. Although Ryff's (1989) purpose in life dimension has relevance to religion and spirituality, how it is conceptualized is different in Eastern and

Middle Eastern belief systems. Additionally, religion and spirituality are central to understanding well-being in these belief systems, whereas this is not the case in Ryff's model.

Given these distinctions, Ryff's (1989) model is probably incompatible with the logic, assumptions, constructs, and hypotheses describing well-being in Eastern and Middle Eastern belief systems. Thus, it may not be applicable to persons who embrace these systems or reside in cultures (countries) where they are prominent. Stated differently, when using Ryff's model and the SPWB to study the well-being of people influenced by these systems it is likely a model equivalence bias would occur reducing the cross-cultural and external validity of the results.

Intervention (Strategy) Bias This is another new form of bias introduced in this article, and also not assessed or understood in a systematic fashion in the past. It is defined as follows: An intervention(s) created in one culture (country) is not applicable to the target culture (country). It also is impossible to use the intervention(s) in the same way, for example, in counseling, consultation, supervision, training, or conflict prevention and/or resolution for specific populations, situations, and/or contexts in a target culture (country). This bias occurs when one or more of the following aspects of the intervention(s) created in one culture (country) are incompatible in the target culture (country): (a) philosophical assumptions, (b) logic, (c) operational definitions, (e) content, (d) target population, context, and/or situation where applied, (e) rationale for using, (f) when it is employed, (g) how it is delivered, (h) parameters of utilization, (i) expected outcomes, and (j) ethical boundaries and constraints. The extent of an intervention bias fluctuates based on the number and to what degree these components are incompatible between the cultures (countries).

There are many examples of interventions derived in the U.S. and employed without adaptation outside of the U.S. that fit the criteria for an intervention bias. One illustration of this bias is Transformative Mediation (Bush & Folger, 1994), a major approach in the U.S. that helps disputants resolve their conflicts by restoring their relationship. This strategy is based on conflict theory, and relational and communication paradigms of conflict and human interactions (Bush & Pope, 2002). The primary goal of this mediation is to help disputants transform their conflict and reconcile their relationship rather than reach an agreement. Transformative mediators must stay neutral to help disputants resolve their issues in their own way, and promote self-determination in the disputants so that they can achieve this outcome (Harper, 2006). The fact the mediator is not directive will introduce an intervention bias if this form of mediation is used with people from some cultures (countries) outside the U.S. For instance, persons from Asia and The Middle East expect a helping professional to be more directive, to possess and display their expertise, and to offer solutions to help them resolve their problems (Hays & Erford, 2018).

Therefore, without modification, relying on a Transformative Mediation approach may seem irrelevant to these individuals. There is no research to support this conclusion. Thus, there is a need to study the appropriateness and effectiveness, or cross-cultural validity of using Transformative Mediation with people from Asia and The Middle East who enter into a mediation process.

Systematic Test of Equivalence Procedure (STEP)

Obviously, depending on the focus of a project, it may be critical to establish equivalence prior to doing a cultural or cross-cultural study to reduce the potential of various kinds biases. In the past, researchers relied most often on non-standardized approaches to establish measurement, linguistic, and/or scalar equivalence before doing a study, and statistics (e.g., factor analysis, confirmatory factor analysis, structural equation modeling; See Ægisdóttir et al., 2008) to explore construct equivalence after collecting the data. Until recently, there was no tool to evaluate the equivalence or cross-cultural validity of a construct or method, including using different research procedures and/or item response formats prior to conducting a study. Gerstein (2018) introduced an approach called the Systematic Test of Equivalence Procedure (STEP) to accomplish these goals.

Initially, STEP was created to offer researchers a concrete strategy for examining the cross-cultural validity of constructs (equivalence and bias) including a new form of equivalence, method equivalence. As stated earlier, this type of equivalence focuses on the relevance of using various research methods and item response formats in different cultures (countries). Previously, STEP involved collecting quantitative relevance ratings and qualitative responses from experts and/or the target participants on the constructs, scale items, and item response formats given the purpose of a study. When performing these tasks, these persons were asked to think about their own culture (country) as they assessed scales and constructs developed in another culture (country). Moreover, to improve the cross-cultural validity and equivalence of a measure these persons were invited to suggest, if they thought necessary, additional new constructs and items tied to these constructs as well as items for the already existing constructs. Descriptive statistics (e.g., frequencies, means, standard deviations) and content analyses were employed when appropriate to evaluate the quantitative and qualitative responses shared by these experts.

In the past, outcomes of a STEP study included: (a) data about construct and method bias and equivalence, (b) content about a relevant method and item response format for a culture (country), and (c) etic and emic constructs and items to conduct cross-culturally valid projects in a target culture (country). To date, STEP studies have investigated using the TNT

(Gerstein et al., 2014; Mayton et al., 1999) in Hong Kong (Gerstein et al., 2018), CVQ (Dik et al., 2012) in Korea (Kim & Gerstein, 2011), the Fear of Rape Scale (Senn & Dzinas, 1996), Frequency of Everyday Harassment Scale (Fairchild & Rudman, 2008), Coping with Harassment Scale (Fairchild & Rudman, 2008), and Objectified Body Consciousness Scale (McKinley & Hyde, 1996) in India (Bellare, & Gerstein, 2018), the Beliefs About Psychological Services Scale (Ægisdóttir, & Gerstein, 2009) in India (Nathani & Gerstein, 2016), Tendency to Seek Help scale (Tinsley, St. Aubin, & Brown, 1982) and Asian American Multidimensional Acculturation Scale (Gim Chung, Kim, & Abreu, 2004) in India (Nathani & Gerstein, 2016), and the Subjective Well-being Assessment Scale (Veronese et al., 2017) with Arab Americans (Atari & Gerstein, 2019).

This article introduces two new applications of STEP designed to investigate (a) theory equivalence and bias, and (b) intervention equivalence and bias. Descriptive statistics (e.g., frequencies, means, standard deviations) and content analyses can be performed when appropriate to assess the data collected when using the two new STEP strategies. When studying the equivalence and bias of a theory to be used in a culture (country) other than the one for which it was originally intended, STEP should be changed to include these four new steps:

1st Step.[1] Each item below is a feature of the ___ (name the theory). Using the scale that follows (1 = Highly Irrelevant to 6 = Highly Relevant), rate the relevance (cultural appropriateness) of the description for a to i with respect to ___ (name the specific phenomena, population, situation, and/or context) keeping in mind the culture (country) of interest:

(a) Philosophical assumptions (include descriptions); (b) constructs (include descriptions); (c) logic (include description); (d) operational definitions (include descriptions); (e) hypotheses (include descriptions); (f) principles or links between the structural association between two or more concepts or constructs (include descriptions); (g) observations (include descriptions); and (i) explanatory power (include descriptions).

2nd Step. If you rated any of the items 1, 2, or 3 in Step 1, please explain your reason(s).

3rd Step. Do you think it is necessary to modify and/or expand the content of any of the items in Step 1 to make them more relevant or culturally appropriate to ___ (name the specific phenomena, population, situation, and/or context) in the target culture (country)? If you do think it is necessary, please explain how you modify and/or expand these?

4th Step. Do you think there are other important features of ___ (name the theory, etc.) not stated in Step 1 that should be included to make ___ (name the theory, etc.) more relevant (culturally appropriate) to ___ (name the specific phenomena, population, situation, and/or context) in the culture (country) of interest? If you think there are other important features, please explain them.

In applying the STEP approach to explore the equivalence of an intervention to be used with specific populations, situations, and/or contexts in the targeted culture (country) other than where it originated, six new steps are involved.

1st Step. Each item below is an aspect of interventions or strategies linked to ___ (name the theory, model, approach, etc.). Using the scale that follows (1 = Highly Irrelevant to 6 = Highly Relevant), rate the relevance (cultural appropriateness) of the description for a to k with respect to ___ (name the specific population, situation, and/or context) keeping in mind the culture (country) of interest:

(a) Philosophical assumptions (include descriptions); (b) logic (include descriptions); (c) operational definitions (include descriptions); (d) content (include descriptions); (e) target population, context, and/or situation where applied (include descriptions): (f) rationale for using (include descriptions); (g) when it is employed (include descriptions); (h) how it is delivered (include descriptions); (i) parameters of utilization (include descriptions); (j) expected outcomes (include descriptions); and (k) ethical boundaries and constraints (include descriptions).

2nd Step. If you rated any of the items 1, 2, or 3 in Step 1, please explain your reason(s).

3rd Step. Each intervention that follows is connected to ___ (name the theory, model, approach, etc.). Rate how relevant (culturally appropriate; 1 = Highly Irrelevant to 6 = Highly Relevant) each is for ___ (name the specific population, situation, and/or context).

List the different interventions and strategies below the rating scale.

4th Step. If you rated any intervention 1, 2, or 3 in Step 3, please explain your reason(s).

5th Step. Do you think it is necessary to modify any intervention listed in Step 3 to make it more relevant or culturally appropriate to be used with ___ (name the specific population, situation, and/or context)? If you think it is necessary, please explain how would you modify each.

6th Step. Do you think there are other interventions that are consistent with ___ (name the theory, model, approach, etc.) that are relevant (culturally appropriate) to utilize with ___ (name the specific population, situation, and/or context)? If you do, what are they?

As the two new adaptations of STEP have yet to be used, it is uncertain if they will be beneficial when testing the equivalence and bias of a theory and intervention. Research is needed, therefore, to establish the usefulness and validity of these

---

[1] A Likert response format was assumed to be a relevant and valid approach to gather responses from the target experts and/or participants. When developing this step for a STEP study, the investigator must first establish the most appropriate response format to use.

adaptations. Until then, it is not possible to draw any firm conclusions about the appropriateness of using these two new STEP applications to study the cross-cultural validity and adaptation of theories and interventions originally created for one culture(s) or country(ies) and potentially to be used in another culture (country).

Conclusion

Clearly, equivalence and bias are critical to performing cross-culturally valid research and practice. Two new forms of equivalence (Theory; Intervention) and bias (Theory; Intervention) were introduced in this article to be used in various scholarly and applied endeavors to account for potential additional components of understanding and demonstrating cross-cultural validity. Moreover, the article introduced an expansion of the STEP method to accommodate investigating the two new types of equivalence and bias. Research studies must be conducted, however, to investigate the validity and relevance of these new forms of equivalence and bias, and applications of STEP.

There is a plethora of research projects that could be conducted to examine the validity of employing STEP to investigate if theories, models, and/or paradigms in psychology, counseling, and education created in one culture (country) are equivalent (bias) or applicable in a target culture (country). For example, the theoretical equivalence of the Emotion Cultivation Process (Wang et al., 2019) model which was created in Taiwan could be researched in other Asian countries and in the West as could Hwang's (1987) theory of Chinese social behavior that is grounded in Confucian ethics and emphasizes social interactions with regard to face and favor. Further, by administering the STEP the theoretical equivalence of the Māori philosophy of health and well-being which is considered unique to New Zealand (Waitoki, Dudgeon, & Nikora, 2018) could be studied with other indigenous populations throughout the world. The theoretical equivalence of principles posited about cognition found in ancient Indian spiritual and philosophical systems also could be examined through the use of the STEP to determine its relevance to modern Western perspectives about cognition (Sedlmeier & Kunchapudi, 2016). Moreover, the theoretical equivalence of Nsamenang's (2001) indigenous West African perspective on human development could be examined in other African nations. Lastly, STEP could be used to investigate the theoretical equivalence or more specifically the operational definitions and hypotheses linked to the Western posttraumatic growth model of Tedeschi and Calhoun (2004) for individuals residing in a diverse array of countries (e.g., Liberia, Egypt, Viet Nam). Performing this project could address a gap in the literature as scholars (Splevins, Cohen, Bowley, & Joseph, 2010; Tennen & Affleck, 2009; Vázquez,

Pérez-Sales, & Ochoa, 2014) have questioned the applicability, validity, and relevance of Tedeschi and Calhoun's (2004) model to non-Western individuals.

There many research and applied projects that also could be performed to explore the validity of utilizing STEP to examine the equivalence or bias of interventions and/or strategies developed in one country (culture) in psychology, counseling, and education to determine if they are applicable in another. For instance, given the zeitgeist of teaching various forms mindfulness in the context of counseling and other psychological delivery systems, it would be revealing to investigate the intervention equivalence of Naikan Therapy, an approach developed in Japan designed to help individuals see themselves with their mind's eye (Tanaka-Matsumi, 2004), for clients residing in other countries. A STEP study also could be conducted to research the intervention equivalence of the WISER (**W**hole School, **I**ndividual, **S**ystem, **E**valuation, & **R**esource Integration; Wang, 2014) model of school counseling which was created in Taiwan for persons living in other countries in and outside of Asia. Additionally, an intervention equivalence project involving an arts-based (i.e., music, poetry) conflict resolution and reconciliation program originally developed in Cyprus (Ungerleider, 1999) could be implemented with persons residing in other parts of the Mediterranean, South America, and the West. One other potential STEP study could examine if intervention equivalence or bias can be demonstrated when West African dances designed to promote peace and community enrichment are employed with non-African populations (Diallo & Hall, 1989). Taken as a whole, the findings of the studies just proposed could provide support for the appropriateness and validity of utilizing STEP to cross-culturally validate intervention approaches.

There are many other empirical and applied projects that could and must be conducted to investigate the equivalence and bias of existing and future theories and interventions. The STEP offers a straightforward and easily employed mixed methodology to accomplish these objectives. Until such efforts are pursued, however, questions remain about the cross-cultural validity, particular in relation to equivalence and bias, of numerous theories and intervention strategies, along with the utility and validity of using STEP to examine these questions.

Compliance with Ethical Standards

References

Allport, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

Atari, R., & Gerstein, L.H. (2019, April). *The assessment of well-being among Arab communities*. Poster presented at the 11th annual Muslim mental health conference. Tempe, AZ.

Bellare, Y., & Gerstein, L.H. (2018, August). *Everyday stranger harassment and coping among Asian Indian and U.S. college students*. Poster presented at Annual Convention of the American Psychological Association, San Francisco, CA.

Bowman, C. G. (1993). Street harassment and the informal ghettoization of women. *Harvard Law Review, 106*, 517–580. https://doi.org/10.2307/1341656.

Bush, R. A. B., & Folger, P. (1994). *The promise of mediation: Responding to conflict through empowerment and recognition*. San Francisco: Jossey-Bass.

Bush, R. A. B., & Pope, S. G. (2002). Changing the quality of conflict interaction: The principles and practice of transformative mediation. *Pepperdine Dispute Resolution Law Journal, 3*(1), 67–96 https://scholarlycommons.law.hofstra.edu/faculty_scholarship/162.

Choi, S., & Choi, S. (2016). Conceptualizing Ryff's psychological well-being model with Confucian perspective. *Korean Journal of Sociology, 50*, 101–123.

Diallo, Y., & Hall, M. (1989). *The healing drum: African wisdom teachings*. Rochester: Destny Books.

Dik, B. J., Eldridge, B. M., Steger, M. F., & Duffy, R. D. (2012). Development and validation of the calling and vocation questionnaire (CVQ) and brief calling scale (BCS). *Journal of Career Assessment, 20*, 242–263.

Fairchild, K., & Rudman, L. A. (2008). Everyday stranger harassment and women's objectification. *Social Justice Research, 21*, 338–357. https://doi.org/10.1007/s11211-008-0073-0.

Fouad, N. A., & Kantamneni, N. (2010). Cultural validity of Holland's theory. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (3rd ed., pp. 703–714). Thousand Oaks, CA: Sage Publications.

Gerstein, L. H. (2011). Research, practice and training across and within cultures: Simply complicated! *Journal of Asia Pacific Counseling, 1*(1), 1–12.

Gerstein, L. H. (2018). Systematic test of equivalence procedure: New method to investigate cross-cultural validity. *Revista de Cercetare si Interventie Sociala, 62*, 278–293.

Gerstein, L. H., Chan, Y., Hutchison, A., Fung, A. L. C., Kinsey, R., & Jeffers, H. (2018). The teenage nonviolence test: Applicability in Hong Kong? *Current Psychology, 37*(1), 313–324. https://doi.org/10.1007/s12144-016-9514-3.

Gerstein, L. H., Heppner, P. P., Ægisdóttir, S., Leung, S. A., & Norsworthy, K. L. (2009). *International handbook of cross-cultural counseling: Cultural assumptions and practices worldwide*. Thousand Oaks, CA: Sage Publications.

Gerstein, L. H., Mayton, D., Hutchison, A., & Kirkpatrick, D. (2014). The teenage nonviolence test: A factor analytic investigation. *Revista de Cercetare si Interventie Sociala, 44*, 9–19.

Gim Chung, R. H., Kim, B. S. K., & Abreu, J. M. (2004). Asian American multidimensional acculturation scale: Development, factor analysis, reliability, and validity. *Cultural Diversity and Ethnic Minority Psychology, 10*, 66–80. https://doi.org/10.1037/1099-9809.10.1.66.

Hardin, E. E., Robitschek, C., Flores, L. Y., Navarro, R. L., & Ashton, M. W. (2014). The cultural lens approach to evaluating cultural validity of psychological theory. *American Psychologist, 69*, 656–668.

Harper, C. (2006). Mediator as peacemaker: The case for activist transformative-narrative mediation. *Journal of Dispute Resolution,* *2*, 595–611 https://scholarship.law.missouri.edu/jdr/vol2006/iss2/10.

Hays, P. A. (2009). Integrating evidence-based practice, cognitive–behavior therapy, and multicultural therapy: Ten steps for culturally competent practice. *Professional Psychology: Research and Practice, 40*, 354–360.

Hays, D. G., & Erford, B. T. (2018). *Developing multicultural counseling competence: A systems approach* (3rd ed.). New York: Pearson.

He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online Readings in Psychology and Culture, Unit 2*. Retrieved from http://scholarworks.gvsu.edu/orpc/vol2/iss2/8

Hwang, K. K. (1987). Face and favor: The Chinese power game. *American Journal of Sociology, 92*, 944–974.

International Test Commission. (2015). *ITC guidelines for translating and adapting tests, version 1*. Retrieved from https://www.intestcom.org/files/guideline_test_adaptation.pdf

Joshanloo, M. (2014). Eastern conceptualizations of happiness: Fundamental differences with western views. *Journal of Happiness Studies, 15*, 475–493. https://doi.org/10.1007/s10902-013-9431-1.

Kim, T.S., & Gerstein, L.H. (2011, August). *Cross-cultural validity of career calling in Korea: Preliminary research*. Poster presented at the annual meeting of the American Psychological Association. Washington, DC.

Lau, W. Y., Chan, C. K., Li, J. C., & Au, T. K. (2010). Effectiveness of group cognitive-behavioral treatment for childhood anxiety in community clinics. *Behaviour Research and Therapy, 48*, 1067–1077. https://doi.org/10.1016/j.brat.2010.07.007.

Leong, T. L., & Brown, M. T. (1995). Theoretical issues in cross-cultural career development: Cultural validity and cultural specificity. In B. W. Walsh & S. H. Osipow (Eds.), *Handbook of vocational psychology: Theory, research, and practice* (2nd ed., pp. 143–180). Mahwah, NJ: Lawrence Erlbaum Associates, Inc..

Lonner, W. J. (1985). Issues in testing and assessment in cross-cultural counseling. *The Counseling Psychologist, 13*, 599–614.

Malpass, R. S., & Poortinga, Y. H. (1986). Strategies for design and analysis. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 47–83). Beverly Hills, CA: Sage Publications.

Mayton, D. M., Weedman, J., Sonnen, J., Grubb, C., & Hirose, M. (1999, August). *The teenage nonviolence test: Internal structure and reliability*. Paper presented at the annual meeting of the American Psychological Association, Boston, MA.

McKinley, N. M., & Hyde, J. S. (1996). The objectified body consciousness scale development and validation. *Psychology of Women Quarterly, 20*, 181–215. https://doi.org/10.1111/j.1471-6402.1996.tb00467.x.

Naeem, F., Phiri, P., Munshi, T., Rathod, S., Ayub, M., Gobbi, M., & Kingdon, D. (2015). Using cognitive behaviour therapy with south Asian Muslims: Findings from the culturally sensitive CBT project. *International Review of Psychiatry, 27*, 233–246. https://doi.org/10.3109/09540261.2015.1067598.

Nathani, P., & Gerstein, L. H. (2016, March). *Acculturation and help seeking: Asian Indians in the United States*. Poster presented at the Great Lakes regional counseling psychology conference, Indiana University, Bloomington, Indiana.

Norsworthy, K. L. (2017). Mindful activism: Embracing the complexities of international border crossings. *American Psychologist, 72*, 1035–1044.

Nsamenang, A. B. (2001). Indigenous view on human development: A West African perspective. In N. J. Smelser & P. B. Baltes (Eds.-in-Chief) (Eds.), International encyclopedia of the social and behavioral sciences. London: Elsevier.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology, 49*, 65–85.

Quintana, S. M., Troyano, N., & Taylor, G. (2001). Cultural validity and inherent challenges in quantitative methods for multicultural research. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (2nd ed., pp. 604–630). Thousand Oaks, CA: Sage Publications.

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*, 1069–1081.

Sedlmeier, P., & Srinivas, K. (2016). How do theories of cognition and consciousness in ancient Indian thought systems relate to current Western theorizing and research? *Frontiers in Psychology, 7*, 343. https://doi.org/10.3389/fpsyg.2016.00343.

Senn, C. Y., & Dzinas, K. (1996). Measuring fear of rape: A new scale. *Canadian Journal of Behavioural Science, 28*, 141–144. https://doi.org/10.1037/0008-400X.28.2.141.

Splevins, K., Cohen, K., Bowley, J., & Joseph, S. (2010). Theories of posttraumatic growth: Cross- cultural perspectives. *Journal of Loss and Trauma, 15*, 259–277.

Tanaka-Matsumi, J. (2004). Japanese forms of psychotherapy: Naikan therapy and morita therapy. In U. P. Gielen, J. M. Fish, M. Jefferson, & J. G. Draguns (Eds.), *Handbook of culture, therapy, and healing* (pp. 277–229). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Tedeschi, R. G., & Calhoun, L. G. (2004). Posttraumatic growth: Conceptual foundations and empirical evidence. *Psychological Inquiry, 15*, 1–18.

Tennen, H., & Affleck, G. (2009). Assessing positive life change: In search of meticulous methods. In C. Park, S. Lechner, A. L. Stanton, & M. H. Antoni (Eds.), *Medical illness and positive life change: Can crisis lead to personal transformation?* (pp. 31–49). Washington, DC: American Psychological Association.

Tinsley, H. E. A., St. Aubin, T. M., & Brown, M. T. (1982). College students' help-seeking preferences. *Journal of Counseling Psychology, 29*, 523–533.

Ungerleider, J. (1999). Music and poetry build bi-communal peace culture in Cyprus: 'My country is cut in two.' *In people building peace: 35 inspiring stories from around the world* (pp. 297–301). Utrecht: European Centre for Conflict Prevention.

Vázquez, C., Pérez-Sales, P., & Ochoa, C. (2014). Posttraumatic growth: Challenges from a cross-cultural viewpoint. In G. A. Fava & C. Ruini (Eds.), *Increasing psychological well-being in clinical and educational settings, cross-cultural advancements in positive psychology* (Vol. 8, pp. 57–74). Dordrecht: Springer Science+ Business Media. https://doi.org/10.1007/978-94-017-8669-0_4.

Veronese, G., Pepe, A., Dagdoukee, J., Addimando, L., & Yaghi, S. (2017). Measuring well- being in Israel and Palestine: The subjective well-being assessment scale. *Psychological Reports, 120*, 1160–1177.

van de Vijver, F. J. R. (2001). The evolution of cross-cultural research methods. In D. Matsumoto (Ed.), *Handbook of culture and psychology* (pp. 77–97). New York: Oxford University Press.

van de Vijver, F. J. R., & He, J. (2016). Bias assessment and prevention in noncognitive outcome measures in context assessments. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning world-wide extended context assessment framework and documentation of questionnaire material* (pp. 229–253). New York: Springer.

Waitoki, W., Dudgeon, P., & Nikora, L. W. (2018). Indigenous psychology in Aotearoa/New Zealand and Australia. In S. Fernando & R. Moodley (Eds.), *Global psychologies mental health and the global south* (pp. 163–184). London: Palgrave Macmillan. https://doi.org/10.1057/978-1-349-95816-0_10.

Wang, L. (2014, July). *The building and implementing of the WISER school counseling model in Taiwan*. Invited speech at the international symposium in school guidance and counseling, Hijiyama University, Hiroshima, Japan.

Wang, L.-F., Wei, M., Koay, E. Y. Y., Lo, M. H., & Lee, M.-Y. (2019). The development and validation of the emotional cultivation scale: An east Asian cultural perspective. *Journal of Counseling Psychology, 66*, 409–423. https://doi.org/10.1037/cou0000346.

Wong, D. F. (2008). Cognitive behavioral treatment groups for people with chronic depression in Hong Kong: A randomized wait-list control design. *Depression and Anxiety, 25*, 142–148.

Ægisdóttir, S., & Gerstein, L. H. (2009). Beliefs about psychological services (BAPS): Development & psychometric properties. *Counselling Psychology Quarterly, 22*, 197–219.

Ægisdóttir, S., Gerstein, L. H., & Cinarbas, D. (2008). Methodological issues in cross-cultural counseling research: Equivalence, bias and translations. *The Counseling Psychologist, 36*, 188–219.