

---

# Equivalence in Research on Positive Development of Minority Children: Methodological Approaches

Fons J.R. van de Vijver and Jia He

---

## Abstract

We describe methodological challenges in studies of positive development of minority youth. A framework is described that captures the main validity threats in this type of research. The framework involves different types of bias (referring to sources of systematic error in studies of minority youth). If there is bias in a study, the comparability of constructs or scores across minority groups can be challenged. Many procedures have been proposed in the literature to deal with such methodological problems, which affect either the design or the analysis of a study. We review such procedures and pay special attention to two topics that attract much attention in recent studies: response styles and mixed methods. It is concluded that sound methods can help to make study results more robust and replicable.

Researchers studying positive development of minority children face many conceptual and methodological challenges (e.g., Motti-Stefanidi 2014; Spencer 1990). We argue that the combination of a culturally sensitive approach and rigorous research methods is crucial in advancing this

research field. Various methodological problems in research on positive development of minority children can be addressed by a careful design and analysis of studies, building on the extensive experience from cross-cultural and developmental psychological studies of the last decades. This chapter focuses on the methodological challenges of comparability and validity in data obtained in different immigrant groups. We first define and review bias and equivalence in cross-cultural research as a theoretical framework to address the comparability and validity issues; we then describe how to deal with bias and equivalence issues, based on empirical examples; we end the chapter with an overview of domains of current research interest and rapid development, and a description of possible future directions.

---

F.J.R. van de Vijver (✉) · J. He  
Department of Culture Studies, Tilburg University,  
P.O. Box 90153, 5000 LE Tilburg, The Netherlands  
e-mail: fons.vandevijver@tilburguniversity.edu

J. He  
e-mail: jamis.he@gmail.com

F.J.R. van de Vijver  
North-West University, Potchefstroom, South Africa

F.J.R. van de Vijver  
University of Queensland, St. Lucia, Australia

## Historical Overview and Theoretical Perspectives

The research on minority children inevitably concerns the appropriateness of measurement and/or the comparability of one minority group to another or to the majority group. According to Poortinga (1995), studies with a cross-cultural comparative nature in the 1940s based their findings on two assumptions: firstly, existing, notably Western, conceptualizations of psychological constructs in one culture are applicable to other cultures; secondly, cultural contexts do not affect the processes and outcomes of an assessment. The second assumption was challenged in the 1960s, and since then culturally sensitive tests were promoted. In the 1980s, the pervasive influences of culture on comparative studies could no longer be ignored, therefore various approaches to adapt tests and sophisticated psychometric analysis tools have been developed (e.g., Cronbach and Drenth 1972; Poortinga and Van de Vijver 1987).

Van de Vijver and Leung (1997) put forward a systematic classification of bias and equivalence, which provides the framework to address methodological issues of cross-cultural studies. *Bias* occurs when score differences on the indicators of a particular construct do not correspond to differences in the underlying trait or ability. It refers to systematic errors in data that are expected to be found again were the study to be repeated and that have an impact on the adequacy of the measure for assessing the purported underlying construct or the average scores of at least of one of the cultures studied. *Equivalence* refers to the implications of bias on test score comparability. It is an indicator of the measurement level at which scores obtained in different cultural groups can be compared.

---

## A Taxonomy of Bias

Three types of bias, namely construct, method, and item bias, are distinguished based on the source of invalidity (Van de Vijver 2015; Van de Vijver and Leung 1997; Van de Vijver and

Tanzer 2004). *Construct bias* occurs when the construct measured is not identical across cultures, either because the concept or because the elements taken to comprise its measure (e.g., attitudes, behaviors, or cognitions) are not comparable (Van de Vijver and Poortinga 1997). For instance, resilience is defined by Ungar (2008) as the capacity of individuals to navigate their way to health-enhancing resources and the capacity of individuals' physical and social ecologies. The definition may well be universally applicable; yet, its manifestations may vary across cultures. With minority children, resilience can manifest itself in different ways. Drop-out is generally recognized as a negative school outcome and an important indicator of poor resilience (e.g., Masten and Coatsworth 1998); however, dropout was found to be a positive indicator of resilience in a group of African Canadian students to establish dignity, personal efficacy, and independence (Dei et al. 1997). A discussion of the pros and cons of both operationalizations is beyond the scope of this chapter. However, the example illustrates that assessing resilience requires researchers to have a good knowledge of the cultural contexts of their studies and to take culture-specific aspects into consideration (e.g., Masten and Motti-Stefanidi 2009).

Another example of construct bias comes from acculturation research. Before describing findings in the acculturation domain, a caveat on terminology is needed. The literature on positive youth development uses the concept of acculturation in two quite distinct meanings. In the first, acculturation is the same as adjustment; well acculturated children refer then to immigrant children who are well adjusted to their new cultural context (e.g., Riggs 2006). The second meaning, adopted here, is broader and views acculturation as orientations towards both the ethnic and mainstream culture (e.g., Neblett et al. 2012). In the psychological acculturation literature the latter view has become dominant (e.g., Sam and Berry 2006), whereas the former view is common in acculturation literature in sociology (Sakamoto et al. 2009) and public health (Lara et al. 2005). The field of positive youth development would gain in clarity if authors were

more explicit about the view on acculturation that they espouse.

It can be argued that these conventional models, notably the view of equating acculturation with adjustment, are increasingly at odds with how acculturation takes place these days. In many contexts youth does not deal with two but with three or even more cultures. An example is “3D-acculturation” which describes how Jamaican immigrants to the US simultaneously negotiate the Jamaican, European American mainstream, and African American cultures (Ferguson et al. 2012). Another example is superdiversity, which refers to neighborhoods where people from many different ethnicities live together (Vertovec 2007); in these neighborhoods acculturation (and identity) processes can no longer be captured in a simple mainstream—immigrant dichotomy but involve multiple allegiances, which can even include cosmopolitanism (as a pan-human identity) (Van de Vijver et al. 2015).

According to Snauwaert et al. (2003), the often employed classification in acculturation orientations (i.e., integration, assimilation, separation, and marginalization; Berry 1997) cannot be taken to refer to immigrant preferences that are the same across all life domains. These authors studied immigrants in Belgium. If acculturation was measured in a contact domain (i.e., perceived desirability of having contacts with both mainstreamers in the country of settlement and immigrants from the same country of origin), integration was preferred; most immigrants find it desirable to have contacts with both groups, which in the conceptual model is taken as evidence in favor of integration. However, if identification with the mainstream and ethnic culture was assessed, Snauwaert and colleagues found support for separation, as identification with the ethnic culture was much stronger than identification with the mainstream culture. These findings are in line with results obtained by Arends-Tóth and Van de Vijver (2003), who found that Turkish-Dutch prefer separation in the private domain and integration in the public domain. These studies suggest that acculturation

orientations vary with personal involvement of the domain and that domains with a strong personal involvement (such as identification and religion) are most resistant to acculturative change. As a consequence, there is no such thing as *the* acculturation orientation of an immigrant or an immigrant group. The common conceptualization of acculturation orientations refers to a domain-independent construct and does not refer to any domain, where such domain dependence may be part and parcel of the construct.

*Method bias* is a generic term for all forms of systematic errors occurring during the process of assessment. It can derive from sampling, structural features of the instrument, or administration processes. *Sample bias* results from incomparability of samples. Cross-cultural variations in sample characteristics can be related to target measures; confounding sample differences could lead to observed score differences in the target measures that do not involve valid cross-cultural differences. A typical case in point is the confounding of educational quality with IQ in cross-cultural comparisons of IQ scores obtained in very different cultural contexts. A meta-analysis revealed that national expenditure on education, which can be taken as a proxy for educational quality, was a predictor of cross-national differences in scores on the Raven’s Progressive Matrices (Brouwers et al. 2009).

*Instrument bias* refers to incomparability arising from instrument characteristics. In comparing the cognitive performances of children from black and white groups in South Africa, Malda et al. (2010) found that children from both groups performed better when the version of the test was designed for their own group, which illustrates how differences in stimulus familiarity, due to cultural differences, can affect cross-cultural comparisons. Another source of instrument bias is response styles, the systematic tendency to use certain answer anchors on some basis other than the target construct (Cronbach 1950). Okanda and Itakura (2010) reported that 3-year-old Japanese children tended to

inappropriately say “yes” to yes–no questions, although they knew the answers to the questions. Comparing different ethnic groups in the Netherlands, He and Van de Vijver (2013) found that Nonwestern immigrants tend to use response moderation strategies such as acquiescent and midpoint responding more than Western immigrants and Dutch mainstreamers.

*Administration bias* can result from different administration conditions (e.g., paper-and-pencil versus online survey, individual versus group administration), unclear instructions, and communication between test administrator and respondents, such as halo effects. For example, African preschoolers showed higher test scores on the Peabody Picture Vocabulary Test-R when tested by African American field staff than by white field staff (Doucette-Gates et al. 1998).

*Item bias*, also known as Differential Item Functioning (DIF), means that an item has a different psychological meaning in the cultures studied. Technically, an item is biased if persons with the same trait or ability, but coming from different cultures, are not equally likely to endorse the item (Van de Vijver and Leung 1997). There are multiple sources of item bias, both linguistic (e.g., poor translation, language features) and cultural (e.g., inapplicability of item contents in different cultures, and items with ambiguous connotations). If item bias is observed, it is important to identify explanations for it (e.g., poor translation or inapplicability of an item in a certain context) (Leung and Van de Vijver 2008). In their study of interethnic attitudes of German mainstream children and Turkish children in Germany, Feddes, Noack, and Rutland (2009) administered a four-item scale, asking how many out-group children were friendly, polite, smart, and bad. For Turkish children, the item “bad” showed DIF, as it caused a lower reliability in the Turkish group. The authors speculated that this might be due to children’s willingness to attribute less positive traits to one group, but not necessarily also attribute more negative traits to this group. Thus, this item was dropped in their further analysis.

## A Taxonomy of Equivalence

Equivalence reflects the level of comparability across cultures. Three levels of equivalence are identified (Van de Vijver and Leung 1997). Whereas bias refers to sources of systematic distortions in cross-cultural comparisons that challenge their validity, equivalence deals with the implications of bias for the comparability of constructs and scores. *Construct equivalence* means that the same theoretical construct is measured in each culture studied. Construct equivalence is a prerequisite for any cross-cultural comparison in any study; without it, no cross-cultural comparison involving the construct would be valid. It is an important first step in the statistical analysis of cross-cultural data to explore the structure of the construct and the adequacy of sampled items. When a construct does not have the same meaning across the cultures, researchers need to acknowledge the incompleteness of conceptualization and can still compare the equivalent facets of the construct (i.e., partial invariance; Byrne et al. 1989). The current Zeitgeist appears to emphasize identity of constructs across cultures. In such a climate the lack of construct equivalence can easily be construed as a reflection of inadequacy of design, sampling, or data administration. This is regrettable as the observation of construct non-equivalence can point to important cross-cultural differences.

*Measurement unit equivalence* (or *metric equivalence*) indicates that measures of interval or ratio level have the same measurement unit (metric) across cultural contexts, but they have different scale origins. When measures show metric equivalence, scores can be compared within cultural groups (e.g., gender differences can be tested in each group), but scores cannot be compared across groups (means of females in one group cannot be compared to means of females in another group).

*Full score equivalence* (or *scalar equivalence*) represents the highest level of equivalence, which means that scales in all groups studied

have the same measurement unit and origin. Observed scores are then free from any type of bias and can be compared directly within and across cultures. When measures show full score equivalence, analyses of variance and *t* tests to examine cross-cultural differences in means are appropriate for (and only for) this level of equivalence.

### **Equivalence as a Characteristic of a Cross-Cultural Comparison**

Equivalence cannot be assumed, but should be empirically demonstrated (Van de Vijver and Poortinga 1997). Consequently, before means are compared across cultures, the first research question to address in minority children studies should be whether there is equivalence, which is the basis for any meaningful and valid conclusion to be drawn (e.g., Benson et al. 2009; Bodkin-Andrews et al. 2010; Buhs et al. 2010). Such an analysis should be routinely conducted in any cross-cultural comparison as a first check, in the same way as internal consistency is reported.

So, cross-cultural studies of immigrant youth should report equivalence and internal consistency. Internal consistency is not an intrinsic characteristic of an instrument, but a characteristic of scores obtained with an instrument in a specific study. The same holds for equivalence. Conclusions about equivalence are based on analyses of data obtained in specific samples. Studies of minority youth are often based on non-probability sampling; as a consequence, generalizability of conclusions about equivalence may be limited. Like internal consistency, equivalence has to be examined and demonstrated in each study.

---

### **Research Measurement and Methodology**

It requires careful design, implementation, and statistical analysis to ensure equivalence (e.g., Cheung and Rensvold 2002; Van de Vijver and

Tanzer 2004). We propose to integrate the strategies at different stages of a study, focusing on best practices in the design and implementation stage and on statistical measures that can empirically test equivalence in the analysis stage.

## **Design and Implementation Strategies**

### **Choice of Instrument**

In the conceptualization and design stage of a study involving minority children, a decision has to be made as to whether an existing instrument will be used or whether a new instrument is to be developed. The choice should depend on more than the availability of an existing instrument. Available instruments have the advantage that they often have been tried and tested, usually in Western groups. Notably when such instruments have shown robust psychometric properties, there may be an expectation that similar characteristics will be found in other cultural groups, although obviously, such characteristics have to be shown. However, it is not a foregone conclusion that instruments applied in their original form can transcend language and culture differences in a new context (e.g., Peña 2007). A major weakness of most existing instruments is that they have not been developed with a cross-cultural target group in mind. Therefore, even if their psychometric properties are favorable, their cultural adequacy may be problematic. It is often all too easily assumed that the instrument may work well and that equivalence analyses can be used to identify possible problems. The development of minority children can be easily construed as “deviant” if majority-based norms are used as starting point (Spencer 1990). Therefore, we argue that the choice of an instrument in studies on minority children and youth should be approached from a broader perspective, balancing substantive and psychometric concentrations.

We argue that three options are available in instrument choice: adoption, adaptation, and assembly (Van de Vijver and Leung 1997). *Adoption* involves the use of the original of a measure (and applying a close translation of an instrument if needed) in another cultural group or

context. The main advantages are that the process is simple to implement and it makes a direct comparison of scores possible (assuming that equivalence can be demonstrated); yet, adoption can only be used when the items in the source and target language versions adequately cover the construct measured and the response formats are appropriate in the new contexts (Harkness 2003). *Adaptation* is a combination of a close translation of certain stimuli and modification of other stimuli; an adaptation is preferred when adoption of all stimuli is inadequate for linguistic, cultural, or psychometric reasons. Nowadays, adaptation has become more and more frequently used and the term is often used as the de facto standard in working with tests in multiple contexts. The change of word signifies the change of emphasis in the process of working with multiple language versions. There is a change from a linguistic to a multidisciplinary perspective in which in addition to language, cultural information and psychological knowledge about the target constructs are viewed as essential in preparing stimulus materials for a new cultural context. *Assembly* refers to the compilation of a new measure; it is indicated when neither adoption nor adaptation would be adequate. Assembly can maximize the cultural appropriateness of an instrument, but it renders numerical comparisons of scores across cultures impossible.

The choice for any of the three options should depend on the target cultures and research aims. Adoption is favored if the goal is to compare scores across-cultures directly, whereas adaptation and assembly are better to enhance the suitability of the instrument for the context in which it will be administered. It is important to note that there is no intrinsically superior option; adoption, adaptation, or assembly can be the best choice, given the balance required between psychometric and cultural considerations.

Adoption has long been viewed as the default choice in cross-cultural research; it is often the “quick-and-dirty” choice that combines relatively little effort to create an instrument for a new group with high levels of comparability. However, adopting an existing instrument may conceal interesting cross-cultural differences that are

not covered by the items of an existing instrument. When adopting an existing instrument, the question is implicitly or explicitly asked whether this instrument, developed for another group, is adequate in the new group. However, the culturally more appropriate question would be: Is the existing instrument the best possible to measure the target construct in the new cultural group? There is a subtle, yet important difference in perspective between these two questions: in the first perspective one culture is taken as frame of reference whereas in the second perspective there is a balance between the perspectives.

### **Pretests and Standardization of Procedures**

It is recommended to carry out pilot studies and cognitive interviews before the field work (Willis 2005), because they can provide information about the adequacy of the instrument in a specific cultural context, reveal possible design problems, and serve to reduce the likelihood of systematic measurement bias. Pilot studies are particularly important when measures are to be assembled from scratch or transported to locations with a large geographic and cultural distance from the culture in which the original instrument was developed. For example, cognitive interviewing elicits respondents’ opinions on the response process, which serves as an effective tool to detect possible bias (e.g., Friborg et al. 2006). The target population should be involved at an early stage for consultation to have assessment of high levels of acceptability and meaningfulness for ethnic minority children (Leff et al. 2006).

When implementing the study, all field workers should abide by a standard protocol, which may include a standardized training for all interviewers. Additional elements aimed to reduce bias, notably method bias, involve the specification of suitable administration conditions (e.g., individual assessment or group assessment) and administration modes (e.g., face-to-face interview, telephone interview, paper-and-pencil survey or online survey) and monitoring the interaction of interviewers and interviewees (in case of halo effects). Other measures such as clear instruction and examples,

detailed field work documentation, assessment of response styles, and test-retest comparisons may also contribute to the minimization of biases (Van de Vijver and Tanzer 2004).

## Statistical Strategies

After administration, various analytic approaches to detect bias and ensure equivalence in collected data can be applied. In this section, we illustrate the utilization of factor analysis at the scale level and Differential Item Functioning analysis (DIF) at the item level.

### Factor Analysis

Both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) can be used to examine construct bias, whereas it is the advantage of CFA that it can also detect item bias (Vandenberg and Lance 2000). When the underlying structure of a construct is unclear, EFA is preferred in investigating and comparing factor structures. The use of EFA (and various other dimensionality-reducing techniques) to study equivalence is based on a simple reasoning: identical constructs are measured in all groups if the structure of an instrument, as analyzed with these techniques, is the same across cultures. So, identity of factors (or dimensions) is taken as sufficient evidence for equivalence. Comparisons of multiple cultures can be conducted either in a pairwise or in a one-to-all manner (in the latter case each culture is compared with the combined solution; Van de Vijver and Poortinga 2002). Target rotations are employed to compare the structure across cultures and to evaluate factor congruence, often by means of the computation of Tucker's phi coefficient, which tests to what extent factors are identical across cultures. Values of the coefficient above 0.90 are usually considered to be adequate and above 0.95 to be excellent (Van de Vijver and Leung 1997). Yağmur and Van de Vijver (2012) compared self-report acculturation and language orientations of Turkish immigrants in four host countries, and the structural equivalence of all the scales were established by

pairwise target comparisons of factor solutions of each scale in the four countries. Such a procedure ensures comparability across Turkish immigrants across host countries.

CFA, as one application of structural equation modeling procedures, is often employed when the structure of the construct can be derived from theory or previous work. An acceptable fit from the CFA indicates that the hypothesized factor structure can be accepted, and there is evidence for equivalence. CFA can test hierarchical models based on information of covariance matrix. For example, if we set to examine whether the same one-factor model holds in various cultures, a series of nested models are usually tested (Cheung and Rensvold 2002). The *configural invariance model* specifies that the same latent construct with the same indicators are assumed. In the *measurement weights model*, factor loadings on the latent variable are constrained to be equal across cultures. If a multigroup confirmatory factor analysis yields a satisfactory fit, the construct under investigation can be said to have construct equivalence. In the *intercept invariance model*, items are constrained to have the same intercept across cultures. A satisfactory fit of the intercept invariance model provides evidence that there is no item bias. Various additional types of invariance have been proposed; for example, in a *structural covariance model* the covariances of latent factors are identical across groups; a *structural residual model* refers to identity of the error component of the latent variable; a *measurement residuals model* specifies the identity of error component of the items. Although it is quite clear that factor loading and intercept invariance are the most important aspects, there is no agreement in the literature about the importance of the other types of invariance. When full invariance cannot be reached, it is also possible to resort to partial invariance by removing the constraints of equal factor loadings and/or intercepts in non-invariant items (Byrne 2001; Byrne and Van de Vijver 2010).

Model fit in CFA is usually evaluated by  $\chi^2$  tests, their significance, and tests of the change in  $\chi^2$  values between different models of invariance. Additional and frequently used indices include the Tucker Lewis Index (TLI; acceptable above

0.90 and excellent above 0.95), Root Mean Square Error of Approximation (RMSEA; acceptable below 0.06 and excellent below 0.04), and Comparative Fit Index (CFI; acceptable above 0.90 and excellent above 0.95). Whether or not a more restricted model is acceptable can be established by the change of Comparative Fit Index; changes of values within 0.01 from the less restricted to the more restricted model usually suggest acceptable fit of the latter (Byrne 2001; Van de Vijver and Poortinga 1997).

The technique, combined with other strategies, has been used in research on positive development of minority children, and proven to be prolific. Michaels, Barr, Roosa, and Knight (2007) used multigroup CFA to check the equivalence of the five domains of self-esteem among Anglo, Mexican, African and Native American youths aged 9–14 years of low-income, inner-city school district in a large metropolitan area in the southwestern United States. Scalar equivalence was reached for the global self-worth and scholastic competence domain, whereas other domains only showed scalar equivalence in some of those groups, indicating that these domains were only meaningful to certain ethnic groups or that items might not adequately represent the construct in all groups. White, Umaña-Taylor, Knight, and Zeiders (2011) investigated the cross-language measurement equivalence of three components of ethnic identity (i.e., exploration, resolution, and affirmation) among Mexican American early adolescents. They reported scalar equivalence of measures of exploration and resolution across language versions and compared full and partial invariance models to draw conclusions on overall comparability. Researchers are encouraged to use these statistical tools to demonstrate comparability of data before making inferences on cultural differences and similarities.

### DIF Analysis

DIF (item bias) analysis targets the identification of anomalous items. DIF refers to the problems caused by the differing probabilities of correctly solving or endorsing an item after matching on the underlying ability that the item is intended to measure in different cultures (Zumbo 2007).

With some exceptions, DIF analysis is applicable only to one-dimensional constructs; therefore for multidimensional constructs, DIF analysis should be performed per dimension. There are many models and procedures one can follow to detect uniform and non-uniform item bias, including ANOVA, logistic regression, item response theory, and the Mantel-Haenszel method. Computer programs to conduct these procedures to study DIF are widely available; examples are logistic regression using SPSS (Zumbo 1999) and Mantel-Haenszel using EASY-DIF (Gonzalez et al. 2011).

Applications of DIF analyses in the literature on positive youth development tend to be part of invariance testing procedures using structural equation modeling where invariance of intercepts (taken as absence of item bias) is tested. A first example comes from a study on invariance in development (Bowers et al. 2010). Using a longitudinal design, these authors tested whether the structure of measures gauging the Five Cs (i.e., Competence, Confidence, Connection, Character, and Caring; see Lerner this volume) of Positive Youth Development were the same among 920 youth across grades 8, 9, and 10. They found evidence for scalar invariance of the measures across these grades, suggesting that their measure can be used to assess the Five Cs in a comparable manner in this age range. A second example is due to Shek and Ma (2010), who tested the gender invariance of the structure of the Chinese Positive Youth Development Scale in a large sample of lower-secondary school students attending a positive youth development program in Hong Kong. They found evidence of scalar invariance of the 15 basic dimensions of this scale and four higher-order factors (i.e., cognitive-behavioral competencies, prosocial attributes, positive identity and general positive youth development qualities).

---

### Focus Areas of Development

In this section we review specific topics in cross-cultural research methods that are relevant for the study of positive youth development. Each

area has the potential to lead to more insights in this development. We discuss (a) response styles and (b) mixed methods. For an overview of recent developments in multilevel modeling, another rapidly evolving field that is relevant for positive youth development, we refer to Asendorpf's chapter in this volume.

## Response Styles

Self-reports using a Likert-type response format continue to be important in the study of youth development. It has been argued repeatedly that their advantages (easy to administer and analyze) are offset by their shortcomings, notably their susceptibility to impression management (Paulhus 1986) and common method variance (Podsakoff et al. 2003). Four response styles have been frequently studied: Acquiescent Response Style (tendency to agree irrespective of item content), Midpoint Response Style (tendency to choose the midpoint or scores around the midpoint of the response scale), Extremity Response Style (tendency to choose the extremes of response scales), and Social Desirability (tendency to choose responses that are in line with perceived norms about what is appropriate in a culture) (e.g., Van Vaerenbergh and Thomas 2013). In one of the few studies that used a measure of response styles (more specifically, social desirability), Papacharisis et al. (2005) evaluated the effectiveness of a life skills program, that was administered to adolescent volleyball and soccer players during their regular practice hours. The trained life skills were goal setting, problem solving, and positive thinking. The authors found that a social desirability scale did not show correlations with questionnaire items, such as items about self-beliefs. The authors concluded that it was very unlikely that social desirability would have any influence on their findings. Gilman et al. (2008) administered the Multidimensional Students' Life Satisfaction Scale to 1338 youth adolescents from Ireland, the US, China, and South Korea. In line with literature on the strong response modesty of East

Asians, the authors found that American and Irish adolescents reported more extremity and acquiescence than Chinese and South Koreans did.

In an attempt to integrate response styles, He and Van de Vijver (2013) found in studies of adults from different ethnic groups in the Netherlands that all styles merge in a single factor; this General Response Style factor has Social Desirability and Extremity Response Style as positive indicators and Acquiescent and Midpoint response Style as negative indicators. At the individual level, the General Response Style is related to all Big Five personality traits and several values (such as embeddedness). At country level, the factor is negatively related to countries' socioeconomic development (with less affluent countries showing higher scores on Social Desirability and Extremity Response Style). The existence of more restrictive norms in less developed countries which emphasize conformity and promote amplified self-expression may underlie these higher scores (He et al. 2014). The General Response Style was even found in a large cross-cultural study that used the Occupational Personality Questionnaire (OPQ32), a forced-choice format personality measure designed to be less affected by response styles than regular personality measures.

Taken together, the evidence suggests that response styles are better viewed as communication styles (amplifying versus moderating of responses), internalized as part of the socialization process (Smith 2004), than as deliberate errors or distortions. Much old research into response styles was based on the idea that these styles should be eliminated, notably the influence of social desirability was to be eliminated (Nederhof 1985). However, there is increasing evidence that validity is not increased by correcting for response styles. Ones et al. (1996) demonstrated that job performance is not better predicted after "peeling off" response styles from applicants' self-reports; in the same vein, He and Van de Vijver (2015) found that statistical corrections for the response style did not affect the size or patterning of cross-cultural differences in

teacher reports. These findings suggest that we may need to reconceptualize and refine our views on response styles.

Most studies of response styles involved adults; therefore, extending these findings to positive youth development awaits confirmation. Still, the picture that emerges is rather clear. Response styles are real, replicable, and can explain sizeable amounts of variance in cross-cultural studies (we found examples of more than 20 %; He et al. 2014); yet, individual differences in response styles explain considerably more variance than cross-cultural differences do. However, statistical corrections may create a false sense of security as these may not increase the validity and typically cannot statistically “explain away” cross-cultural differences.

## Mixed Methods

Scientific progress can be stifled by persistent controversies. The best best-known example in the field of cross-cultural methods is the emic-etic distinction (Pike 1967). The emic perspective is associated with the qualitative approach (understanding a culture from within), whereas the etic perspective is associated with the quantitative approach (comparing samples from different cultural groups). The two camps have long been at loggerheads. Yet, there is good reason for trying to integrate qualitative and quantitative procedures more (Van de Vijver 2015); the strengths and weaknesses of both procedures are complementary, so that they do not only have their own methods but also their own research questions. The richness of qualitative research, with its emphasis on an open approach to reality, has its main strength in exploring new constructs and cultures. The main strength of quantitative procedures is their rigor and allowance to test specific hypotheses. So, qualitative procedures are best in the context of discovery, whereas quantitative procedures are best in the context of justification (Reichenbach 1938).

In the last decades we have witnessed the emergence of so-called mixed methods that combine qualitative and quantitative methods

(Tashakkori and Teddlie 2010). The most frequent combination is a study in which in a first phase qualitative methods are used (in cross-cultural studies this is often used to examine the context), followed by a quantitative stage in which a survey is conducted. However, other combinations are possible, such as a quantitative study with a qualitative follow-up (the procedure is described by Onwuegbuzie and Leech 2004). The statistical procedure will yield outliers, which would be adolescents with exceptionally low or high resilience scores, given their parenting style scores. Follow-up interviews with these adolescents are then conducted to identify which factors could have contributed to their extreme scores.

An important and not yet fully developed methodological component of mixed-methods is triangulation (Denzin 2012), which amounts to the question of how the qualitative and quantitative evidence can be combined. If two types of evidence provide convergent information, triangulation is straightforward. As an example, Van de Vijver et al. (2015; see also Blommaert 2013) were interested in the identity of immigrants in a superdiverse area in Oud-Berchem, a suburb of Antwerp, Belgium. Superdiversity refers to the presence of many ethnic groups in a single neighborhood, thereby creating their own mixtures, dynamics, and relationships. The common distinction between ethnic and mainstream culture does not suffice to describe the cultural richness and complexities of such neighborhoods. Using an ethnographic approach, these authors found a rather strong cohesion in the area despite its huge ethnic diversity. This qualitative leg of the study led to the expectation that the immigrant inhabitants would show rather strong Belgian, ethnic, and cosmopolitan identities, which was confirmed in a quantitative survey. The convergence of the qualitative and quantitative results made the results easy to interpret. Suppose now that Belgian and cosmopolitan identity scores would have been low. Triangulation of results could then become problematic unless a clear interpretation of the low scores could be given (e.g., poor measurement or complete lack of coherence in the neighborhood).

Mixed-methods applications in the field of positive youth development have been reported. For example, Henderson et al. (2005) were interested in the influence of organized camp programs on growth and development of youth in the US. They derived quantitative (pre-post surveys) and qualitative evidence (observations) from a total of six camps. The data included pre- and post-questionnaires given to campers (youth) to measure domains, such as positive identity, social skills, positive values, and thinking and physical skills. The qualitative part focused more on camp characteristics and included observations on the structure and delivery of the program. There was some convergence of the main findings of both approaches: the camps that showed significant pre-post differences had also the programs that yielded more favorable qualitative data. Yet, at a more detailed level, it was difficult to link qualitative data about the camps to (quantitative) changes in youth. The latter is a common problem in triangulating quantitative and qualitative data: both types of data often address somewhat different issues (such as a more contextual, qualitative analysis and a more individual-oriented quantitative approach).

A second example uses a very different and common type of triangulation: qualitative evidence is converted to quantitative evidence (or the other way around) so that triangulation takes place within a single data mode. This is easier than cross-mode triangulation. For example, if qualitative data are quantified, regular statistical approaches can be employed to analyze convergence with the other, quantitative data. In a study designed to explore links between perceived family support, acculturation, and life satisfaction, Edwards and Lopez (2006) studied Mexican American adolescents. Qualitative data came from a thematic analysis of open-ended responses to a question about life satisfaction; notably if existing instruments may fail to cover all relevant aspects in a certain group, such an open approach has important advantages. The other constructs were assessed using quantitative instruments. The quantified life satisfaction data were then used as dependent variables in a regression analysis, with perceived support from family and Mexican and

Anglo acculturation orientations as predictors. As expected, both independent variables were significant predictors of life satisfaction.

---

## Future Directions

Positive development of minority children is an emerging field; its potential to further promote children's welfare is remarkable, as it has been repeatedly demonstrated that protective factors are at least as important than risk factors in child development (e.g., Motti-Stefanidi et al. 2012). Compared to decades ago, an impressive number of studies have been conducted and we have gained valuable experience informing us what (not) to do in these studies. We argued in this chapter that the quality of research on minority children could be improved by paying more attention to methodological issues. Adequately designed, conducted, and analyzed studies are often easier to interpret, have to deal with fewer alternative score interpretations, and are more insightful as they deal with cultural factors more adequately. If we use the tools and experience reviewed in this chapter, the future of positive development studies on minority children is bright and we can expect to considerably enlarge our insights in the cross-cultural differences and similarities of child development.

---

## References

- Arends-Tóth, J. V., & Van de Vijver, F. J. R. (2003). Multiculturalism and acculturation: Views of Dutch and Turkish-Dutch. *European Journal of Social Psychology*, 33, 249–266. doi:10.1002/ejsp.143
- Benson, N., Oakland, T., & Shermis, M. (2009). Cross-national invariance of children's temperament. *Journal of Psychoeducational Assessment*, 27, 3–16. doi:10.1177/0734282908318563
- Berry, J. W. (1997). Immigration, acculturation and adaptation. *Applied Psychology: An International Review*, 46, 5–68.
- Blommaert, J. (2013). *Ethnography, superdiversity and linguistic landscapes: Chronicles of complexity*. Bristol, United Kingdom: Multilingual Matters.
- Bodkin-Andrews, G. H., Haa, M. T., Cravena, R. G., & Yeunga, A. S. (2010). Factorial invariance testing and latent mean differences for the self-description

- questionnaire II (Short Version) with indigenous and non-indigenous Australian secondary school students. *International Journal of Testing*, 10, 47–79. doi:[10.1080/15305050903352065](https://doi.org/10.1080/15305050903352065)
- Bowers, E. P., Li, Y., Kiely, M. K., Brittan, A., Lerner, J. V., & Lerner, R. M. (2010). The five Cs model of positive youth development: A longitudinal analysis of confirmatory factor structure and measurement invariance. *Journal of Youth and Adolescence*, 39, 720–735. doi:[10.1007/s10964-010-9530-9](https://doi.org/10.1007/s10964-010-9530-9)
- Brouwers, S. A., Van de Vijver, F. J. R., & Van Hemert, D. A. (2009). Variation in Raven's progressive matrices scores across time and place. *Learning and Individual Differences*, 19, 330–338. doi:[10.1016/j.lindif.2008.10.006](https://doi.org/10.1016/j.lindif.2008.10.006)
- Buhs, E. S., McGinley, M., & Toland, M. D. (2010). Overt and relational victimization in Latinos and European Americans: Measurement equivalence across ethnicity, gender, and grade level in early adolescent groups. *Journal of Early Adolescence*, 30, 171–197. doi:[10.1177/0272431609350923](https://doi.org/10.1177/0272431609350923)
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Byrne, B. M., & Van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10, 107–132. doi:[10.1080/15305051003637306](https://doi.org/10.1080/15305051003637306)
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105, 456–466. doi:[10.1037/0033-2909.105.3.456](https://doi.org/10.1037/0033-2909.105.3.456)
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233–255. doi:[10.1207/s15328007sem0902\\_5](https://doi.org/10.1207/s15328007sem0902_5)
- Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and Psychological Measurement*, 10, 3–31.
- Cronbach, L. J., & Drenth, P. J. D. (Eds.). (1972). *Mental tests and cultural adaptation*. The Hague, the Netherlands: Mouton.
- Dei, G. J. S., Mazzuca, J., McIsaac, E., & Zine, J. (1997). *Reconstructing "drop-out": A critical ethnography of the dynamics of Black students' disengagement from school*. Toronto, Canada: University of Toronto Press.
- Denzin, N. K. (2012). Triangulation 2.0. *Journal of Mixed Methods Research*, 6, 80–88. doi:[10.1177/1558689812437186](https://doi.org/10.1177/1558689812437186)
- Doucette-Gates, A., Brooks-Gunn, J., & Chase-Lansdale, P. L. (1998). The role of bias and equivalence in the study of race, class and ethnicity. In V. C. McLoyd & L. Steinberg (Eds.), *Studying minority adolescents: Conceptual, methodological, and theoretical issues* (pp. 211–236). Mahwah, NJ: Erlbaum.
- Edwards, L. M., & Lopez, S. J. (2006). Perceived family support, acculturation, and life satisfaction in Mexican American youth: A mixed-methods exploration. *Journal of Counseling Psychology*, 53, 279–287. doi:[10.1037/0022-0167.53.3.279](https://doi.org/10.1037/0022-0167.53.3.279)
- Feddes, A. R., Noack, P., & Rutland, A. (2009). Direct and extended friendship effects on minority and majority children's interethnic attitudes: A longitudinal study. *Child Development*, 80, 377–390. doi:[10.1111/j.1467-8624.2009.01266.x](https://doi.org/10.1111/j.1467-8624.2009.01266.x)
- Ferguson, G. M., Bornstein, M. H., & Pottinger, A. M. (2012). Tridimensional acculturation and adaptation among Jamaican adolescent-mother dyads in the United States. *Child Development*, 83, 1486–1493. doi:[10.1111/j.1467-8624.2012.01787.x](https://doi.org/10.1111/j.1467-8624.2012.01787.x)
- Friborg, O., Martinussen, M., & Rosenvinge, J. H. (2006). Likert-based vs. semantic differential-based scorings of positive psychological constructs: A psychometric comparison of two versions of a scale measuring resilience. *Personality and Individual Differences*, 40, 873–884. doi:[10.1016/j.paid.2005.08.015](https://doi.org/10.1016/j.paid.2005.08.015)
- Gilman, R., Huebner, E. S., Tian, L., Park, N., O'Byrne, J., Schiff, M., et al. (2008). Cross-national adolescent multidimensional life satisfaction reports: Analyses of mean scores and response style differences. *Journal of Youth and Adolescence*, 37, 142–154. doi:[10.1007/s10964-007-9172-8](https://doi.org/10.1007/s10964-007-9172-8)
- Gonzalez, A., Padilla, J.-L., Hidalgo, M. D., Gomer-Benito, J., & Benitez, I. (2011). EASYDIF: Software for analysing differential item functioning using the Mantel-Haenszel and standardization procedures. *Applied Psychological Measurement*, 35, 490–499. doi:[10.1177/0146621610381489](https://doi.org/10.1177/0146621610381489)
- Harkness, J. A. (2003). Questionnaire translation. In J. A. Harkness, F. J. R. Van de Vijver, & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 19–34). New York, NY: Wiley.
- He, J., & Van de Vijver, F. J. R. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Differences*, 55, 794–800. doi:[10.1016/j.paid.2013.06.017](https://doi.org/10.1016/j.paid.2013.06.017)
- He, J., & Van de Vijver, F. J. R. (2015). Effects of a general response style on cross-cultural comparisons: Evidence from the Teaching and Learning International Survey. *Public Opinion Quarterly*, 79, 267–290. doi:[10.1093/poq/nfv006](https://doi.org/10.1093/poq/nfv006)
- He, J., Bartram, D., Inceoglu, I., & Van de Vijver, F. J. R. (2014). Response styles and personality traits: A multilevel analysis. *Journal of Cross-Cultural Psychology*, 45, 1028–1045. doi:[10.1177/0022022114534773](https://doi.org/10.1177/0022022114534773)
- Henderson, K. A., Powell, G. M., & Scanlin, M. M. (2005). Observing outcomes in youth development: An analysis of mixed methods. *Journal of Park and Recreation Administration*, 23, 58–77.
- Lara, M., Gamboa, C., Kahramanian, M. I., Morales, L. S., & Hayes Bautista, D. E. (2005). Acculturation and Latino health in the United States: A review of the literature and its sociopolitical context. *Annual Review of Public Health*, 26, 367–397. doi:[10.1146/annurev.publhealth.26.021304.144615](https://doi.org/10.1146/annurev.publhealth.26.021304.144615)
- Leff, S. S., Crick, N. R., Angelucci, J., Haye, K., Jawad, A. F., Grossman, M., et al. (2006). Social cognition in

- context: Validating a cartoon-based attributional measure for urban girls. *Child Development*, 77, 1351–1358. doi:[10.1111/j.1467-8624.2006.00939.x](https://doi.org/10.1111/j.1467-8624.2006.00939.x)
- Leung, K., & Van de Vijver, F. J. R. (2008). Strategies for strengthening causal inferences in cross cultural research: The consilience approach. *International Journal of Cross Cultural Management*, 8, 145–169. doi:[10.1177/1470595808091788](https://doi.org/10.1177/1470595808091788)
- Malda, M., Van de Vijver, F. J. R., & Temane, M. (2010). Rugby versus soccer in South Africa: Content familiarity contributes to cross-cultural differences in cognitive test scores. *Intelligence*, 38, 82–595. doi:[10.1016/j.intell.2010.07.004](https://doi.org/10.1016/j.intell.2010.07.004)
- Masten, A. S., & Coatsworth, J. D. (1998). The development of competence in favorable and unfavorable environments: Lessons from research on successful children. *American Psychologist*, 53, 205–220. doi:[10.1037/0003-066X.53.2.205](https://doi.org/10.1037/0003-066X.53.2.205)
- Masten, A. S., & Motti-Stefanidi, F. (2009). Understanding and promoting resilience in children: Promotive and protective processes in schools. In T. R. Gutkin & C. Reynolds (Eds.), *The handbook of school psychology* (4th ed., pp. 721–738). Hoboken, NJ: Wiley.
- Michaels, M. L., Barr, A., Roosa, M. W., & Knight, G. P. (2007). Self-esteem: Assessing measurement equivalence in a multiethnic sample of youth. *Journal of Early Adolescence*, 27, 269–295. doi:[10.1177/0272431607302009](https://doi.org/10.1177/0272431607302009)
- Motti-Stefanidi, F., Asendorpf, J. B., & Masten, A. S. (2012). The adaptation and well-being of adolescent immigrants in Greek schools: A multilevel, longitudinal study of risks and resources. *Development and Psychopathology*, 24, 451. doi: [10.1017/S0954579412000090](https://doi.org/10.1017/S0954579412000090)
- Motti-Stefanidi, F. (2014). Immigrant youth adaptation in the Greek school context: A risk and resilience developmental perspective. *Child Development Perspectives*, 8, 180–185. doi:[10.1111/cdep.12081](https://doi.org/10.1111/cdep.12081)
- Neblett, E. W., Rivas-Drake, D., & Umaña-Taylor, A. J. (2012). The promise of racial and ethnic protective factors in promoting ethnic minority youth development. *Child Development Perspectives*, 6, 295–303. doi:[10.1111/j.1750-8606.2012.00239.x](https://doi.org/10.1111/j.1750-8606.2012.00239.x)
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15, 263–280.
- Okanda, M., & Itakura, S. (2010). When do children exhibit a “yes” bias? *Child Development*, 81, 568–580. doi:[10.1111/j.1467-8624.2009.01416.x](https://doi.org/10.1111/j.1467-8624.2009.01416.x)
- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology*, 81, 660–679. doi:[10.1037/0021-9010.81.6.660](https://doi.org/10.1037/0021-9010.81.6.660)
- Onwuegbuzie, A. J., & Leech, N. L. (2004). Enhancing the interpretation of “significant” findings: The role of mixed methods research. *The Qualitative Report*, 9, 770–792.
- Papacharisis, V., Goudas, M., Danish, S. J., & Theodorakis, Y. (2005). The effectiveness of teaching a life skills program in a sport context. *Journal of Applied Sport Psychology*, 17, 247–254. doi:[10.1080/10413200591010139](https://doi.org/10.1080/10413200591010139)
- Paulhus, D. L. (1986). Self-deception and impression management in test responses. In A. Angleitner & J. S. Wiggins (Eds.), *Personality assessment via questionnaire* (pp. 143–165). New York, NY: Springer-Verlag.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, 78, 1255–1264. doi:[10.1111/j.1467-8624.2007.01064.x](https://doi.org/10.1111/j.1467-8624.2007.01064.x)
- Pike, K. L. (1967). *Language in relation to a unified theory of structure of human behavior* (2nd ed.). The Hague, the Netherlands: Mouton Press.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879–903. doi:[10.1037/0021-9010.88.5.879](https://doi.org/10.1037/0021-9010.88.5.879)
- Poortinga, Y. H. (1995). Cultural bias in assessment: Historical and thematic issues. *European Journal of Psychological Assessment*, 11, 140–146. doi:[10.1027/1015-5759.11.3.140](https://doi.org/10.1027/1015-5759.11.3.140)
- Poortinga, Y. H., & Van de Vijver, F. J. R. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology*, 18, 259–282. doi:[10.1177/0022002187018003001](https://doi.org/10.1177/0022002187018003001)
- Reichenbach, H. (1938). *Experience and prediction*. Chicago, IL: University of Chicago Press.
- Riggs, N. R. (2006). After-school program attendance and the social development of rural Latino children of immigrant families. *Journal of Community Psychology*, 34, 75–87. doi:[10.1002/jccp.20084](https://doi.org/10.1002/jccp.20084)
- Sakamoto, A., Goyette, K. A., & Kim, C. (2009). Socioeconomic attainments of Asian Americans. *Annual Review of Sociology*, 35, 255–276. doi:[10.1146/annurev-soc-070308-115958](https://doi.org/10.1146/annurev-soc-070308-115958)
- Sam, D. L., & Berry, J. W. (Eds.). (2006). *The Cambridge handbook of acculturation psychology*. Cambridge, United Kingdom: Cambridge University Press.
- Shek, D. T., & Ma, C. M. (2010). Dimensionality of the Chinese positive youth development scale: Confirmatory factor analyses. *Social Indicators Research*, 98, 41–59. doi:[10.1007/s11205-009-9515-9](https://doi.org/10.1007/s11205-009-9515-9)
- Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35, 50–61. doi:[10.1177/0022022103260380](https://doi.org/10.1177/0022022103260380)
- Snauwaert, B., Soenens, B., Vanbeselaere, N., & Boen, F. (2003). When integration does not necessarily imply integration: Different conceptualizations of acculturation orientations lead to different classifications. *Journal of Cross-Cultural Psychology*, 34, 231–239. doi:[10.1177/0022022102250250](https://doi.org/10.1177/0022022102250250)
- Spencer, M. B. (1990). Development of minority children: An introduction. *Child Development*, 61, 267–269. doi:[10.1111/j.1467-8624.1990.tb02778.x](https://doi.org/10.1111/j.1467-8624.1990.tb02778.x)
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage handbook on mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage Publications.

- Ungar, M. (2008). Resilience across cultures. *British Journal of Social Work*, 38, 218–235. doi:[10.1093/bjsw/bc1343](https://doi.org/10.1093/bjsw/bc1343)
- Van de Vijver, F. J. R., & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology*, 33, 141–156. doi: [10.1177/0022022102033002002](https://doi.org/10.1177/0022022102033002002)
- Van de Vijver, F. J. R. (2015). Methodological aspects of cross-cultural research. In M. Gelfand, Y. Hong, & C. Y. Chiu (Eds.), *Handbook of advances in culture & psychology* (Vol. 5, pp. 101–160). New York, NY: Oxford University Press.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis of comparative research*. Thousand Oaks, CA: Sage.
- Van de Vijver, F. J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 29–37. doi:[10.1027/1015-5759.13.1.29](https://doi.org/10.1027/1015-5759.13.1.29)
- Van de Vijver, F. J. R., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, 54, 119–135. doi:[10.1016/j.erap.2003.12.004](https://doi.org/10.1016/j.erap.2003.12.004)
- Van de Vijver, F. J. R., Blommaert, J. M. E., Gkoumasi, G., & Stogianni, M. (2015). On the need to broaden the concept of ethnic identity. *International Journal of Intercultural Relations*, 46, 36–46. doi:[10.1016/j.ijintrel.2015.03.021](https://doi.org/10.1016/j.ijintrel.2015.03.021)
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25, 195–217. doi:[10.1093/ijpor/eds021](https://doi.org/10.1093/ijpor/eds021)
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70. doi:[10.1177/109442810031002](https://doi.org/10.1177/109442810031002)
- Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, 30, 1024–1054. doi:[10.1080/01419870701599465](https://doi.org/10.1080/01419870701599465)
- White, R. M. B., Umaña-Taylor, A. J., Knight, G. P., & Zeiders, K. H. (2011). Language measurement equivalence of the ethnic identity scale with Mexican American early adolescents. *The Journal of Early Adolescence*, 31, 817–852. doi:[10.1177/0272431610376246](https://doi.org/10.1177/0272431610376246)
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage.
- Yağmur, K., & Van de Vijver, F. J. R. (2012). Acculturation and language orientations of Turkish immigrants in Australia, France, Germany, and the Netherlands. *Journal of Cross-Cultural Psychology*, 43, 1110–1130. doi:[10.1177/0022022111420145](https://doi.org/10.1177/0022022111420145)
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-Type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4, 223–233. doi:[10.1080/15434300701375832](https://doi.org/10.1080/15434300701375832)