Methodological aspects of cross-cultural developmental studies

van de Vijver, F.J.R.; Hofer, J.; Chasiotis, A.

Published in:
Handbook of cross-cultural developmental science

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

2
Methodology

FONS J. R. VAN DE VIJVER, JAN HOFER, and ATHANASIOS CHASIOTIS

Introduction

There is a growing interest in the study of cultural factors in developmental science. It is easy to see why. Understanding development requires the delineation of both universal and culture-specific variations in processes and outcomes. Cross-cultural studies have clearly shown that we cannot assume that findings arrived at in Western societies have universal validity. Un÷iversality and culture specificity are testable claims rather than assumptions; moreover, we know from existing cross-cultural studies that methodological aspects require much attention because we can take less for granted in cross-cultural studies than in monocultural studies. For example, instruments that have shown good reliability and validity in Western cultures may lose these properties in a non-Western context. Cross-cultural developmental studies have yielded various interesting results. We present two examples.

Research indicates that the adverse academic effects of authoritarian parenting found in Western countries may not be universal. Chao (1994; Bornstein and Lansford, Chapter 14, this volume; Steinberg, Lamborn, Dornbusch, and Darling, 1992) administered questionnaires of parental control and authoritative-authoritarian parenting style and Chinese child-rearing items involving the concept of "training" (hard work, self-discipline, and obedience) to Chinese American and European American mothers of preschool-aged children. The Chinese American mothers were found to score significantly higher on authoritarian parenting style and training ideologies. In a second study by the same author, parenting styles and school performance of European American adolescents and first- and second-generation Chinese Americans were compared. A positive association between authoritative parenting and school performance was found for the European Americans and, to a lesser extent, for second-generation Chinese Americans, but not for first-generation Chinese Americans (Chao, 2001). Baumrind's (1967) distinction among authoritarian, authoritative, and permissive parenting may need conceptual elaboration if it is to be used in non-Western contexts. AQ1

A second example comes from a study on short-term memory span in Libyan children (Shebani, Van de Vijver, and Poortinga, 2005). Baddeley (1997) formulated the phonological loop hypothesis, which holds that memory traces decay rapidly unless refreshed by rehearsal. The hypothesis predicts that people have a longer memory span for shorter stimuli. In Arabic, each digit can be pronounced in two ways that differ in length (short form and long form). Libyan boys and girls of two grades were presented either the short or long form of digits in recall and pronunciation tasks. Rehearsal speed (a measure of refreshment rate) was positively related with memory span, and children showed a longer memory

span for shorter stimuli, thereby confirming the validity of the phonological loop model. The Arabic language provides a context to test Baddeley's model that cannot be achieved in other languages.

The goal of this chapter is to provide an overview and illustration of the major methodological aspects of cross-cultural studies in developmental science. The chapter comprises four parts. The first part describes bias and equivalence of measurements. In the second part, the theoretical background on bias and equivalence is further elaborated, and then methodological implications of conceptual issues of defining of culture, sampling of cultures, and descriptions of developmental contexts are addressed. The third part describes methodological and statistical tools that hold important promise for enhancing the quality of cross-cultural developmental studies. Multilevel models, integrative research designs combining qualitative and quantitative data, and natural experiments are presented as examples. Conclusions are drawn in the final part; it is argued that to advance our level of knowledge, cross-cultural developmental studies should attempt to integrate conceptual models and advanced methodological and statistical tools and move beyond the dichotomy of qualitative and quantitative approaches.

Key Issues in the Methodology of Cross-Cultural Developmental Studies

Bias and Equivalence

Cross-cultural developmental studies require data from different groups. Once we have collected data from different contexts, we can compare data from various groups and examine cultural differences or similarities across groups. Are such comparisons valid? More than 30 years ago, Triandis (1976) noted that research may become increasingly complex when we depart from the neat designs of experimental psychology with their tight control of ambient variables. The questions of to what extent measurements are equally appropriate for each of the groups under investigation and whether observations and test scores can be interpreted in the same way across populations are particularly relevant in cross-cultural psychology (Van de Vijver and Tanzer, 2004).

Widely used psychological theories and constructs have been developed predominantly in Western contexts. Cross-cultural research is indispensable to evaluate the generalizability of these theories or constructs. In other cultures, other constructs may be important that have never been instantiated in standard Western instruments because they are only locally relevant or have been overlooked in the West (Winter, 1996; Zhang and Bond, 1998). In hindsight, various historical examples of generalizations about differences in traits and abilities of cultural groups can be seen as based on psychometrically poor measures (Van de Vijver and Tanzer, 2004). It is crucial in cross-cultural research to address the equivalence of measurements and test bias because cross-cultural measurements may be distorted by various factors (Van de Vijver and Leung, 1997). For example, test–retest studies of cognitive instruments have shown that persons with little previous test experience often show considerable score gains at retesting and that retesting increases the predictive validity of an instrument in such a population (e.g., Nkaya, Huteau, and Bonnet, 1994). Retest score gains that differ across cultures indicate that the scores at the first occasion were not fully comparable across these cultures; the score gains may be due to memory effects, a better understanding of the test instructions, or the lower novelty of the testing situation so that participants feel more comfortable (Van de Vijver, Daal, and Van Zonneveld, 1986). Without retest data, the nature and size of the cross-cultural differences could be easily misinterpreted. An evaluation of cross-cultural findings without any concern for the comparability of the findings is risky (Dana, 2000). The computation of cross-cultural differences in *t* tests or analyses of variance without examining the comparability of the findings can easily lead to incorrect conclusions. We explain later how comparability can be evaluated.

AQ2

Three hierarchically linked levels of equivalence are commonly distinguished in the literature: construct (structural and functional) equivalence, measurement unit equivalence, and scalar (full score) equivalence (Van de Vijver and Leung, 1997; see also Poortinga, 1989). The term *bias* is generally used to describe "nuisance" factors that negatively affect the equivalence of measurements across different (cultural) groups. Concepts of equivalence and bias do not refer to intrinsic properties of an instrument but rather to characteristics of a given comparison of test scores between cultural groups. Van de Vijver and Leung (1997) described three major types of bias—construct bias, method bias, and item bias—depending on whether the comparability is challenged by the construct, the administration method or samples to measure the construct, or specific items.

Equivalence of Construct Construct equivalence is present when the same construct is measured across cultural groups (regardless of whether measurement procedures are identical in each cultural group). Nomological networks of the instruments in cultures at hand can be examined to demonstrate equivalence of constructs. Functional equivalence of constructs is observed when similar patterns of convergent and discriminant relations with theoretically relevant variables are found across groups. In contrast, construct inequivalence or bias is present when respondents from different cultural groups do not ascribe the same meaning to the construct as a whole or if there is only partial overlap in the construct's definition across cultures.

Cross-cultural studies of achievement motivation provide a good example of construct bias. In Western studies, the need for achievement is typically defined as an individualistic desire to do things well and to overcome obstacles (McClelland, 1985). McClelland and colleagues were criticized for neglecting contextual and cultural determinants of achievement motivation. In line with such arguments, a number of studies point to a qualitative difference in achievement AQ3
motivation in non-Western societies that is characterized by a pronounced social-oriented element (e.g., Doi, 1982; Kagan and Knight, 1981). In particular, scholars studying Chinese culture emphasized that pushing oneself ahead of others and actively striving toward self-enhancement are not universally valued (Bond, 1986; Yu, 1996). Rather, the concept of a social-oriented achievement motive reflects a need to meet expectations of significant persons and groups (e.g., family and peers). Winter (1996) argued that a kind of mastery motive (a general desire for agency and control) is probably an evolved innate aspect of our biological heritage; still, cultural specificities in childrearing practices, socialization patterns, dominant religious belief systems, values, and social rules to sanction individuals' behavior (Keller and Greenfield, 2000) will involve distinct experiences of rewards and punishments. These differences in cultural practices will eventually lead to the development of differences in terms of concerns for achievement, releasing stimuli, domains of action, and evaluation standards (Phalet and Lens, 1995). Consequently, a monocultural approach based on a Western conception of achievement does not cover all relevant aspects of the construct in non-Western cultures.

Another example may be taken from cross-cultural research on theory of mind. A basic assumption of mainstream developmental science is that everyday knowledge of human psychology is the same everywhere. This universality claim for mentalistic understanding and its development ("theory of mind") (Premack and Woodruff, 1978) has important implications for cultural and interpersonal understanding. If the conviction that other humans are mental beings whose ways of behavior are based on certain states of mind (needs, beliefs, or emotions) holds true, we also tend to view mind as rational and able to control emotions, intentions, and thereby actions. However, there are also reasons to assume culture-specific conceptualizations of mind. There might be cultures that explain actions by referring less to inner mental states and more to contextual factors or even to spirits outside the body (Lillard, 1998). In a review discussing cultural variations in theory

of mind, Lillard (1998) claimed that the European American model of folk psychology is not universal.

A way to answer the question of universality of the concept of folk psychology is to consider its development. Chasiotis, Kiessling, Hofer, and Campos (2006) investigated the relation of theory of mind (measured here as false-belief understanding) and inhibitory control (the ability to suppress a reaction and activate another). The latter is assumed to be an important prerequisite of the former (compare Chasiotis, Kiessling, Winter, and Hofer, 2006). Three samples of preschoolers from Europe (Germany), Africa (Cameroon), and Latin America (Costa Rica) were involved. After controlling for age, gender, siblings, language understanding, and mother's education, culture did not have a moderating effect; each culture showed the same relation between conflict inhibition and false-belief understanding. Furthermore, delay inhibition was not a significant predictor of false-belief understanding in any culture. These results are in line with studies involving American or Asian samples (Carlson and Moses, 2001; Sabbagh, Xu, Carlson, Moses, and Lee, 2006), indicating the possible universality of the relation between delay inhibition and false-belief understanding. Cameroonian children scored significantly lower in theory of mind than the other two cultures; they also showed lower scores in conflict inhibition and higher scores in delay inhibition. The differences in mean scores make the culture-invariant relation between conflict inhibition and false-belief understanding even more interesting because the mean differences are observed against a backdrop of culture-invariant relations between the concepts. These findings suggest that the interdependent parenting goals of obedience and compliance might be related to better delay inhibitory performance and lower false-belief understanding in children (Chasiotis, Bender, Kiessling, and Hofer, in press).

Equivalence of Measurement Unit The second level of equivalence is called measurement unit equivalence. It is present when measures have the same unit of measurement across cultures but have different origins. A difference in origin might emerge when sources of method bias shift mean scores in at least one of the cultures. Depending on its source, it is useful to differentiate three types of method bias, namely administration bias, sample bias, and instrument bias.

Administration bias is caused by sources associated with the particular form of test administration. For example, differences in physical and technical environmental administration conditions, such as noisy versus quiet test locations or the presence of unfamiliar measurement devices (e.g., tape recorder or video camera), and differences in social environmental conditions, such as individual versus group administration and amount of space between participants, may cause substantial cross-cultural differences in target variables (e.g., test performance) and various nontarget variables (e.g., willingness to self-disclose). Further examples of administration bias are ambiguous instructions for study participants and/or guidelines for administrators, communication problems between respondents and administrators (e.g., language problems and violation of cultural communication norms), or the obtrusiveness of the mere presence of a person from a different culture (Super, 1983).

Sample bias occurs when cultural samples are not comparable with respect to relevant background characteristics other than the target construct. As a consequence, observed cross-cultural differences may reflect the target construct but may also be attributed to the influence of "nuisance variables" (e.g., level of education and volunteer bias). For example, in research on theory of mind, mothers' educational level and/or socioeconomic status are predictors of the children's understanding of false-belief tasks (Cole and Mitchell, 2000). Thus, it is essential to carefully balance cultural samples early in the recruitment process.

Finally, instrument bias reflects instrument characteristics causing cross-cultural differences that are unrelated to the target construct. The most important bias that leads to differences in origins of an instrument is group differences in familiarity with test material (e.g., items and

response procedures) and response styles (e.g., acquiescence, extremity ratings, and social desirability). Different familiarity with measurements is a recurrent problem in cross-cultural studies, especially if the study involves "remote" cultural samples. Deregowski and Serpell (1971) found differences in performance between Scottish and Zambian children in sorting photographs but not in sorting miniature models. To reduce group differences in familiarity with stimulus material and testing, Hofer and colleagues (Hofer and Chasiotis, 2004; Hofer, Chasiotis, Friedlmeier, Busch, and Campos, 2005) adapted test instructions for picture-story tests because people from non-Western cultures were more likely to produce mere descriptions of picture cards rather than to create fantasy stories. By giving participants from all cultural groups a detailed and vivid introduction to the picture-story test, such group differences were minimized. Probably the most studied sources of instrumental bias have been cultural differences in response styles (e.g., Marín, Gamba, and Marín, 1992; Van Hemert, Van de Vijver, Poortinga, and Georgas, 2002). Participants with a higher age, lower education, and lower socioeconomic status are more likely to show acquiescence and social desirability (Grimm and Church, 1999; Van de Vijver and Leung, 2001).

Full Score Equivalence The third level of equivalence, namely scalar or full score equivalence, is present when the measurement has the same measurement unit and origin across cultures. This level of equivalence is needed for direct cross-cultural comparisons of means, such as in *t* tests and analyses of variance. A source of bias that may obstruct reaching this level of equivalence (in addition to the presence of construct bias or method bias) is called item bias or differential item functioning (Holland and Wainer, 1993). Item bias is based on characteristics of single items (e.g., nonequivalent content or wording). An item is taken to be biased when people with the same underlying psychological construct (e.g., achievement motivation) from different cultural groups respond diversely to a given item (e.g., test item or picture card). The problem of item bias has often been studied for educational and cognitive tests, has been less studied for self-report measurements such as personality scales, and has been largely neglected for other types of measurements such as projective measurements (Hofer et al., 2005; Van de Vijver, 2000).

Item bias is often caused by a poor translation or adaptation of items. Although translations are linguistically correct, the item may still not be suitable for use across cultures due to culture-bound connotations or linguistic idiosyncrasies (Van de Vijver and Tanzer, 2004). In some cases, items that are useful in one culture do not make sense or are inappropriate in another culture. For example, "I make all my own clothes and shoes" and "I have attended school at some time during my life" (taken from the Personality Research Form; Jackson, 1984) may be useful items to assess a careful and purposeful pattern of responding among Western participants. However, one can easily imagine cultural contexts where such items lose their intended meaning. Comparing the stimulus material used for the assessment of implicit motives among German and Zambian adolescents, Hofer and Chasiotis (2004) found that picture cards clearly differed in their strength to trigger motive imagery across cultural samples. One of the cards depicted a white-collar employee in an office with a family picture at his desk. Stories by German participants were scored much higher for need for affiliation, whereas stories written by Zambian respondents were scored higher for achievement motive.

How Can We Identify and Remedy Various Sources of Bias?

Numerous strategies are described in the literature to identify and remedy the three types of bias. Two main approaches have been proposed to detect biased items: the judgmental approach and the statistical approach. In judgmental procedures, inappropriate items are identified by cultural experts. Few studies have applied this approach. The majority of studies examine item bias by employing different statistical methods depending on the measurement level of items,

number of (cultural) groups, or sample size (for an overview, see Van de Vijver and Leung, 1997; Van de Vijver and Tanzer, 2004). Despite the many statistical techniques available and the numerous studies conducted, our knowledge about factors that induce item bias is limited. It is often difficult to find convergence between judgmental and statistical approaches (e.g., Engelhard, Hansche, and Rutledge, 1990). No specific item features have been found to increase or decrease item bias. Therefore, it is recommended to combine both judgmental and statistical strategies in research. Cultural experts may initially scrutinize wording and content of items, and statistical procedures are used for bias examination in a second step.

To minimize or measure the influence of method bias, various steps can be taken in the design and implementation of a cross-cultural study, such as an intensive training of test administrators; detailed instructions and manuals for administration, scoring, and interpretation; and balancing samples with respect to important participant and context variables. Furthermore, test–retest designs and an examination of response styles may obviate the risk of method bias.

Both design- and analysis-oriented ways of addressing construct bias have been proposed. A combination of the two kinds of procedures is recommended. Various statistical techniques are available to identify construct bias that usually amount to a comparison of data structures across cultural groups, such as the comparison of factor structures (see Van de Vijver and
AQ4 Leung, 1997). One could avoid bias in cross-cultural research by developing culture-specific, indigenous measurements. This procedure might be particularly applicable when there are serious doubts about the expected equivalence or the universal nature of the construct under investigation (Church, 2001). For example, indigenous research on personality in China has provided evidence for the existence of an additional dimension beyond the Five-Factor Model, labeled *Interpersonal Relatedness* (Cheung, 2006). If the research focus is more on universal features and on developing instruments that are applicable across cultures, cultural decentering may be an adequate procedure to avoid construct bias. This procedure involves a simultaneous development of the instrument in several cultures accompanied by a gradual adaptation of the measure, such as elimination of culture-specific words and concepts (e.g., Tanzer, Gittler, and Ellis, 1995). An alternative is the convergence approach, which involves independent measurement development in different cultures and a subsequent employment of all measures in all cultural samples under investigation (see Campbell, 1986).

In conclusion, meaningful comparisons between cross-cultural groups can only be made if sources of bias are addressed and successfully ruled out. Neglecting issues of equivalence in cross-cultural research leads to interpretation problems because alternative explanations, such as differences in construct definition or response styles, cannot be ruled out. Thus, an integrated examination of construct, method, and item bias is highly desirable to enhance our understanding of cultural differences and universals.

AQ5 *How to Approach Culture?*

There are two different traditions in defining culture in cross-cultural psychology (Goodnow, Chapter 1, this volume; Lonner and Adamopoulos, 1997; Rohner, 1984; Segall, 1984). The first views culture as a *molar* Gestalt consisting of interrelated parts. Psychological phenomena are inextricably linked to their cultural context. Culture and psyche are said to make up each other;
AQ6 an essential feature of culture is shared meaning, which is created in the process of interactions and communications among a culture's members. Negotiation between cultural members leads to shared meaning and intersubjectivity. This view is commonly found in cultural psychology (Greenfield, 1997; Miller, 1997). The emphasis on the interrelations of cultural elements is often based on the view that culture as a concept has a limited dimensionality. The best known example is the popular dimension of individualism–collectivism (e.g., Triandis, 1995). The dimension refers to how the relation to the individual and the group is viewed in a culture (Greenfield,

2000). Individualistic societies prioritize individuals and emphasize their independence and uniqueness, whereas collectivistic societies prioritize the group (particularly in-groups, such as the family) by emphasizing the relatedness of individuals. This difference has numerous ramifications for psychological functioning and the way in which a society is organized. For example, socialization practices can be seen as functional adaptations that prepare children for a more individualistic or more collectivistic lifestyle. There is evidence that mother–child interactions vary as a function of individualism–collectivism (Keller, Yovsi, et al., 2004). Mothers in collectivistic societies tend to emphasize relatedness more, whereas mothers in more individualistic societies put more emphasis on autonomy. This difference in emphasis starts when children are very young.

The second view on culture is more *molecular*. Culture is seen as a set of antecedent variables that are linked with psychological functioning in feedback loops (Poortinga and Van de Vijver, 1987). Studies in this tradition typically attempt to identify specific cultural factors that can account for psychological outcomes. A well-known example is the study by Segall, Campbell, and Herskovits (1966) on illusion susceptibility. They argue in their "carpentered world hypothesis" that living in a Western society where geometric shapes, such as trade lines, rectangles, straight lines, and square corners, abound affects susceptibility to some visual illusions, such as the Müller-Lyer illusion. Westerners are more susceptible to these illusions than non-Westerners. Westerners are inclined to apply perceptual habits (interpreting three-dimensional cues to two-dimensional pictures) that are functionally adaptive in daily life but that are maladaptive in the perception of illusion figures.

The literature has long been dominated by the view that molar and molecular conceptions of cultures are incompatible. The two views were even associated with different methodologies. The molar tradition was more associated with ethnographic and qualitative means of data collection and analysis ("cultural psychology"), and the molecular tradition was more associated with the comparative, quantitative tradition ("cross-cultural psychology"). Increasingly, investigators acknowledge that both approaches have their merits and shortcomings and should be seen as complementary (instead of incompatible). A study of the relation between parenting style and children's autonomy could be carried it out in a single country to see whether culture-specific aspects of the concepts and relations can be identified; alternatively, the relation could also be studied in a comparative perspective. The methodology that can be employed will largely depend on the availability and desirability to use standardized instruments. The use of such instruments is not recommended in a monocultural study that attempts to unravel culture-specific features, whereas their use is much more likely and desirable in a cross-cultural study. There is a growing rapprochement between the approaches and appreciation of the complementary nature of molar and molecular models and methods.

Description of context Comparisons are only possible with a common point of reference. One commonly used point of reference in developmental studies of behavior is defined by universal developmental tasks (Keller, 2007). Because enculturation is co-constructed through participation in cultural practices during everyday activities (Rogoff, 2003), behavioral expressions of these tasks are embedded in their cultural context. Keller and her collaborators (Keller, 2007) have documented systematic differences in cultural models of parenting defined by broader cultural models of the self. Two contrasting prototypes can be identified: a model of interdependence, which is more adaptive in subsistence-based, less affluent families with low education and early reproduction, and a model of independence, which is more adaptive in "Western," more affluent urban areas where parents have a higher education and reproduce late. Moreover, variations of these two cultural dimensions of independence and interdependence can be postulated (Kagitcibasi, 2005) and empirically verified (e.g., Keller, Yovsi, et al., 2004). An autonomous-related sociocultural orientation has been found to prevail in urban middle-class families in

AQ7

traditionally interdependent societies, such as in Costa Rica, China, and India (Kagitcibasi, 2005; Keller, 2007).

Because of these variations in sociocultural orientation, the cultural context of investigation can vary starting from the participation procedure (e.g., who decides about participation), the assessment situation (e.g., what do the participants expect from the research), or defined communication styles (e.g., politeness norms of visiting families and required unobtrusiveness of the researcher). Most important, for the urban Western context, common scenarios of mother–child interactions, like a free-play situation between mother and child, might not be equally familiar or accepted in rural or tribal contexts such as India or Cameroon (cf. Keller, 2007). Interview studies can also be problematic because of different cultural conventions pertinent to interview situation, such as who is allowed to provide what kind of information. Such problems can only be treated with a culturally informed methodology, preferably by combining qualitative and quantitative approaches.

Sampling of Cultures There are essentially three ways in which cultures are sampled in developmental studies. The first and most common is *convenience sampling*. Cultures are selected because of availability, easy access, networks of researchers from the countries involved, or some other reason not related to substantive research questions. Such comparisons were common and relevant in the first generation of cross-cultural studies. Those studies helped to set up an empirical database mapping cross-cultural similarities and differences; however, both the quality and quantity of comparative studies have increased so much in the last decades that convenience sampling is now often seen as problematic. First, it is often difficult to link cultural factors to observed differences in psychological variables without a theory to sample cultures. Second, decades of cross-cultural research have shown that convenience sampling leads to biased sampling. Meta-analyses of cross-cultural studies indicate that a few geographical areas dominate the cross-cultural literature; examples are North America, East Asia, and Western Europe. Areas with very different cultures, such as Africa and South America, are much less represented in the literature (Öngel and Smith, 1994; Smith, Harb, Lonner, and Van de Vijver, 2001).

In *systematic (or theory-guided) sampling*, cultures are selected on theoretical grounds. Berry (1976) was interested in field dependence (independence), which is the tendency to be more (or less) influenced in the perception of an object by its background. It was hypothesized that agricultural societies that are more focused on collectivism and conformity encourage their members to be less autonomous and hence can be expected to show a higher level of field dependence. Two types of cultural groups (Canadian hunters-gatherers and African agriculturists) were selected to evaluate this hypothesis. The main strength of systematic sampling is its theoretical basis. Cross-cultural differences that are based on systematic sampling are easier to interpret than differences found in studies using convenience sampling; systematic sampling makes it easier to rule out more alternative interpretations of the cross-cultural differences observed. The systematic sampling of cultures can also show some methodological weaknesses, in particular when only a few cultures are considered. Campbell (1986) has repeatedly argued that two-culture studies are often difficult to interpret because of the many rival explanations that can be put forward; studies involving more than two cultures are less prone to rival alternative explanations. The argument also pertains to studies using systematic sampling strategies. Berry's (1976) work involved a comparison of Canadian hunters-gatherers and African agriculturists. When the study was replicated in Central Africa with culturally similar groups, the findings only partially supported the original hypothesis (Berry et al., 1986).

Finally, in *random sampling*, a probability sample of cultures is drawn. This sampling frame is used for mapping cross-cultural differences and evaluating the universality of the structure of a construct (structural equivalence) or the accuracy of a pan-cultural theory. Because of practical

constraints, it is almost impossible to obtain a truly random sample; however, samples of large-scale studies may approximate a probability sample. Recent examples of large-scale studies can be found in personality (McCrae et al., 2005), social psychology (Schwartz, 1992), organizational psychology (House et al., 2004; Smith, Peterson, and Schwartz, 2002), and survey research (Inglehart, 1997). Large-scale studies in the developmental area always involve comparisons of school performance and educational achievement. A good example is the Programme for International Student Assessment (PISA), which was initiated by the Organisation for Economic Development and Cooperation (OEDC; 2003). Another example is the Trends in International Mathematics and Science Study 2003 (TIMSS), which was organized by the International Association for the Evaluation of Educational Achievement (Mullis, Martin, and Foy, 2005). Both projects involve more than 40 countries and aim at providing policymakers with international benchmarks for identifying the strengths and weaknesses of various educational systems. Despite the impressive size of these studies, the cultural variability of the participating countries is limited, with an overrepresentation of affluent countries and an underrepresentation of developing countries. As a consequence, these studies of educational achievement do not provide a truly universal picture but may well provide a random sample of affluent countries.

Culture and Data Analysis There are various ways to approach culture in comparative designs. AQ8
The distinction between molar and molecular approaches to culture can be used to describe the decisions to be made. In data analyses using a molar approach, there is a tendency to treat culture (or cultural syndromes such as individualism–collectivism) as a nominal variable and to contrast cultures, thereby examining the range of influence of culture in psychological functioning. These studies often have an implicit focus on finding cross-cultural differences. Studies using a molecular approach typically do not start from cultural syndromes but from more specific cultural factors, such as socialization practices and schooling quality. AQ9

Culture plays a slightly different role in the analyses of both approaches. A molar approach takes culture as a starting point and addresses psychological consequences of culture (e.g., Which developmental milestones are affected by a culture's level of individualism?). A molecular approach attempts to decompose culture by unpackaging it (Whiting, 1976). Observing a cross-cultural difference in some psychological process is the beginning rather than the endpoint of a study. Cross-cultural studies are more successful if they can explain more observed cross-cultural differences in psychological function. In statistical terms, the explanatory variables are used as covariates in an analysis of covariance or as independent variables in a hierarchical regression analysis. The analysis addresses the question of to what extent observed cross-cultural differences can be "explained away" by the explanatory variables (Poortinga and Van de Vijver, 1987; Van de Vijver and Leung, 1997). In an analysis of variance with culture as the independent variable and psychological scores as dependent variables, the significance and effect size of culture indicate how much cross-cultural variation there is to be explained; after correction for covariates, the same analysis of variance, now using the residual scores as dependent variables, indicates how much cross-cultural variation is still left. The more cross-cultural variation that is left, the less successful our explanatory variables have been.

Thus, the seemingly paradoxical consequence of analyses of this kind is that we want to get rid of culture as an explanatory variable in cross-cultural research and identify contextual variables that are held responsible for sample differences across cultures. As an example, Chasiotis, Hofer, and Campos (2006) first regressed implicit parenting motivation on the variable "younger siblings." In the next step, the unstandardized residual of implicit parenting motivation of that regression analysis was re-entered in an analysis of variance with culture as predictor. The analysis of variance with the residual of implicit parenting motivation as the dependent variable and culture as the predictor showed a remarkable decrease in effect size of culture from .050 to .041,

which means that 18% of the impact of culture on implicit parenting motivation was caused by the existence of younger siblings. The psychologically rather crude measure of "number of siblings" reduced the effect from a medium (.050) to a small (.019) size, meaning that 62% of the original effect size of culture on implicit parenting motivation could be traced back to sibling effects.

Promising Avenues

In this section, we describe three methodological developments that hold potential for further integrating cultural factors in developmental studies: multilevel designs and multilevel models, integrative approaches, and natural experiments.

Multilevel Designs and Multilevel Models Recent developments in statistics have made it possible to address variation in nested structures. For example, children are nested in families, which are nested in cultures. Multilevel studies consider variation at two or more levels concurrently, such as individual and cultural levels. Two kinds of multilevel approaches have been developed (Hox, 2002; Muthén, 1994; Raudenbush and Bryk, 2002). The first addresses the structural equivalence of concepts at different levels of aggregation. McCrae et al. (2005) were interested in the question of whether the five-factorial structure of personality that is found at the individual level would also be observed at country level. After aggregating their individual-level data (N = 12,156) at country level (N = 51), the authors found the same structure as commonly observed at individual level. This support for the structural equivalence of personality at the two levels implies that individual and country differences in personality scale scores have the same meaning. Similarity of meaning is not a foregone conclusion. It could well be that method bias (e.g., response style differences or incomparable samples) induces a change of meaning after aggregation. Shen and Pedulla (2000; see also Stanat and Luedtke, in press) analyzed data from TIMSS 2003. The authors examined the relation between self-reported mathematics ability and actual mathematics performance. The relation was studied both at individual level per country and at country level. At the individual level, the findings revealed a positive relation (the values of the correlation ranged from r = .12 to r = .47 across the participating countries). However, the country-level correlation was negative, r = –.57. The authors attributed the reversal of the correlation to cross-cultural differences in self-evaluations of ability. Scale scores at the country level reflect not only self-evaluations of ability, but also the tendency of cultural groups to be self-critical or modest. There is evidence to the effect that persons from East Asian cultures do not display the self-presentation styles of Westerners and show a modesty bias (Fahr, Dobbins, and Cheng, 1991; Shikanai, 1978; Takata, 1987).

The second type of multilevel model addresses the interplay of levels. These models address the question of to what extent a phenomenon at a certain level (e.g., the reading achievement of a child) is associated with variables at different levels (e.g., intelligence and socioeconomic status at individual level, school quality at school level, and educational expenditure at country level). Most examples come from the educational domain. Van Langen, Bosker, and Dekkers (2006; see also Stanat and Luedtke, in press) examined performance gaps between boys and girls in the 42 countries participating in the PISA project. Data were analyzed at individual, school, and country level. Student achievements in mathematics, science, and reading were predicted on the basis of individual-level characteristics (e.g., gender and socioeconomic status), school characteristics (e.g., mean socioeconomic status, gender composition, and public versus private school types), and country characteristics (e.g., mean socioeconomic status, female economic activity rate, and gender empowerment index). The analysis of reading test scores revealed a significant interaction between gender and economic activity rates of women; mathematics and science achievement did not show this expected interaction. The reading performance gap in favor of girls tends to be larger in countries with higher female economic activity rates.

Integrative Approaches The second important methodological avenue for developmental studies is the use of integrative approaches that combine input from different methods, cultures, and/or ages (Bornstein, 2002). An example is the cross-cultural study that uses "method triangulation" (Keller, 2007, p. 57); interviews and verbal material of observed interactions are used as qualitative methods, and a quantitative methodology is used in the analysis of questionnaires and videotaped or in situ spot observations of behavior. The goal of the inductive and recursive qualitative codings, namely to gather instances for further examination, is more pragmatic, and the quantitative methodology allows the analytical testing of hypotheses generated by qualitative means. The qualitative methodology can also be used to substantiate and differentiate quantitative results (Georgas, Berry, Van de Vijver, Kagitcibasi, and Poortinga, 2006; Keller, Hentschel, et al., 2004). As another example, Bornstein et al. (2004) asked mothers of 20-month-olds in Argentina, Belgium, France, Israel, Italy, the Republic of Korea, and the United States to fill out comparable vocabulary checklists for their children. In each language, children's vocabularies contained relatively more nouns than other word classes, such as verbs and adjectives. Furthermore, the authors provide a brief description of the main features of the languages. This (qualitative) description is used to provide the linguistic context against which the universally high prevalence of nouns can be interpreted.

Another integrative approach can be found in psychometrically sound cross-cultural applications of implicit measures on life satisfaction (Hofer, Chasiotis, and Campos, 2006), generativity (Hofer, Busch, Chasiotis, Kärtner, and Campos, 2008), and parenthood (Chasiotis, Hofer, et al., 2006). As an example of a multimethod integrative design, Hofer et al. (2006) replicated earlier findings in monocultural studies with German (Brunstein, Schultheiss, and Grässmann, 1998) and Zambian adolescents (Hofer and Chasiotis, 2003) in a cross-cultural study among Germans, Costa Ricans, and Cameroonians using bias-free implicit and explicit measures of relatedness as predictors of life satisfaction. As an explicit measure, the Benevolence Scale of the Schwartz Value Survey was used; and as an implicit measure, a bias-free Thematic Apperception Test–type picture-story test measuring the need for affiliation-intimacy was administered. Results revealed AQ10
that an alignment of implicit motives and self-attributed values was associated with an enhanced life satisfaction across cultures. Chasiotis, Hofer, et al. (2006) assessed explicit and implicit motivation for parenthood combined with a cross-cultural developmental perspective. They assumed that childhood context is important for the emergence of caregiving motivation. A model was tested across cultures in which being exposed to interactive experiences with younger siblings in childhood elicits nurturant implicit motivations that, in turn, lead to more conscious feelings of love toward children in adulthood, which are linked to parenthood. The path model describing this developmental pathway was valid in male and female participants and in all cultures under examination. This study supported the view that childhood context variables such as birth order might exert similar influences on psychological, somatic, and reproductive trajectories across different cultures (see also Chasiotis, Keller, and Scheffer, 2003).

Natural Experiments The last promising area involves the use of *natural experiments* (Scheier, 1959). The large-scale natural experiment of the division of Germany provided an opportunity to compare the influence of four decades of different sociopolitical structures in the former East and West Germany, which were culturally largely similar before the country was split at the end of World War II (Noack, Hofer, Kracke, and Klein-Allermann, 1995). Chasiotis, Scheffer, Restemeier, and Keller (1998) compared two similar urban areas in East (Halle) and West Germany (Osnabrück). Mother–daughter dyads from West and East Germany were analyzed to test the assumption that the onset of puberty is a context-sensitive marker of a reproductive strategy by comparing female parental and filial childhood context and somatic development in both regions. The effect of two different conditions of childhood context continuity on daughter's

age at menarche was tested with the maternal age at menarche controlled. Linear regression models showed that mother's age at menarche only predicted the daughter's age at menarche if the childhood contexts of the mother's and daughter's generations were similar, which was only the case in the West German sample. In East Germany, the mother's age at menarche had no significant effect, and the variance of daughter's age at menarche was explained by filial childhood context variables alone. The comparison of the two samples of mother–daughter dyads in Eastern and Western Germany demonstrated the context sensitivity of somatic development and also showed that this context sensitivity is in line with the evolutionary theory of socialization: What seems to be inherited is not the timing of puberty per se, but the sensitivity for the prepubertal childhood context.

Another example concerns schooling. The relevance of schooling in cognitive development has been discussed for a long time. The Russian cultural-historical school argued that the skill to read and write has a formative influence on abstract thinking (Tulviste, 1991). The problem with testing this position is that reading and writing are acquired in the school context; therefore, schooling and the skills to read and write are confounded in nearly all populations. The confounding does not exist among the Vai in Liberia, where indigenous script is taught by adults to children in an informal setting. The Vai culture provides a natural experiment to avoid this confounding. Scribner and Cole (1981) compared the cognitive test performance of Vai illiterate adults without schooling, literate adults without schooling, and literate adults who were formally schooled. Literates in Vai outperformed illiterates only on tasks that required skills that are also used in dealing with specific Vai script features. High levels of specificity in differences between schooled and unschooled literates of an indigenous script were replicated among the Cree in Canada by Berry and Bennett (1991). Schooling affords children with tangible gains in development that typically focus on their efficient problem-solving strategies and not on their overall level of cognitive functioning (Case, Demetriou, Platsidou, and Kazi, 2001; Cole and Cagigas, Chapter 6, this volume; Schliemann, Carraher, and Ceci, 1997).

Studies of the relation between schooling and cognitive development that are conducted among children suffer from confounding chronological and educational age. The strong correlation of both kinds of ages in countries with compulsory schooling makes it impossible to estimate their relative contribution to cognitive development. The educational system among the Kharwar in India provides a natural experiment to overcome this confounding (Brouwers, Mishra, and Van de Vijver, 2006). The sample comprised 201 schooled and unschooled children from 6 to 9 years of age. The test battery contained various cognitive tests that used either a formal (school-related) or local stimulus content. Confirmatory factor analyses supported similar hierarchical factor structures, with general intelligence in the apex, for both unschooled and schooled children. The per annum score increments of chronological age were approximately twice as large as those of educational age. The study pointed to the important role of everyday experiences in the development of basic features of cognitive functioning.

These examples show how natural experiments can provide important insights by unconfounding variables that co-occur in most cultures. However, such experiments also have limitations. The most salient is the impossibility to manipulate the natural conditions. For example, the finding by Brouwers et al. (2006) that chronological age has more influence on cognitive test scores than educational age has to be interpreted against the backdrop of an overall low quality of schooling among the Kharwar. It was impossible to contrast good and bad schools in the area because of a lack of quality differentiation among the schools.

Conclusion

Developmental science assumes a multidisciplinary vantage point to understand ontogenetic development. Factoring culture into the equation is essential for a comprehensive understanding

of this development. A few models integrate individual- and culture-level perspectives on development, such as Bronfenbrenner's (1977) ecological model, Super and Harkness's (1986) ecological niche model, and Cole's (1999) culture-context model. These models provide important first steps; yet, their heuristic value is limited. For example, studies of the ecological niche are often aimed at merely demonstrating the existence of cross-level relations (Van de Vijver and Poortinga, 2002). Recent methodological and statistical advances, such as multilevel models and analyses of bias and equivalence, enable a more fine-grained analysis of interactions at different levels. It is important to use these tools at a larger scale; yet, the use of more sophisticated research designs and statistical techniques alone is unlikely to generate new insights. In our view, it is important to integrate theory, design, and analysis as much as possible so as to enhance study quality. Theoretical sophistication and methodological sophistication are sometimes seen as incompatible; relatively few studies combine both types of sophistication. Cross-cultural studies deepen our understanding of the cultural factor in development. This goal is more likely to be achieved if we combine a theoretical framework that captures the interaction of individual and cultural factors, such as the three models mentioned earlier, with a sophisticated design and a data analysis that can model the interactions studied. Furthermore, it is important to include relevant contextual data in our studies, either quantitative (as part of the statistical analyses) or qualitative (as a description of the cultural context of the study). Numerous methodological and statistical procedures described in the current chapter, such as analyses of bias and equivalence, multimethod approaches, multilevel models, and natural experiments, are useful tools to increase the quality of our studies and the validity of our findings. If we are successful in integrating theoretical and methodological innovation, developmental cross-cultural studies have a bright future. It is easy to recognize that cultural factors are important in understanding developmental processes and outcomes. However, to be successful, we need to move beyond this recognition; we need to generate knowledge that is relevant for developmental science in general. The current chapter is intended to show how sophisticated methodological tools in cross-cultural developmental studies can contribute to generate new knowledge and advance development science.

AQ11

References

Baddeley, A. (1997). *Human memory: Theory and practice* (revised). Mahwah, NJ: Erlbaum.

Baumrind, D. (1967). Child care practices anteceding three patterns of preschool behavior. *Genetic Psychology Monographs, 75,* 43–88.

Berry, J. W. (1976). *Human ecology and cognitive style. Comparative studies in cultural and psychological adaptation*. Beverly Hills, CA: Sage.

Berry, J. W., and Bennett, J. A. (1991). *Cree syllabic literacy: Cultural context and psychological consequences*. (Cross-Cultural Psychology Monographs No. 1). Tilburg, The Netherlands: Tilburg University Press.

Berry, J. W., Van de Koppel, J. M. H., Sénéchal, C., Annis, R. C., Bahuchet, S., Cavalli-Sforza, L. L., et al. (1986). *On the edge of the forest: Cultural adaptation and cognitive development in Central Africa*. Lisse, The Netherlands: Swets and Zeitlinger.

Bond, M. H. (1986). The social psychology of Chinese people. In M. H. Bond (Ed.), *The psychology of Chinese people* (pp. 213–264). Hong Kong: Oxford University Press.

Bornstein, M. H. (2002). Toward a multiculture, multiage, multimethod science. *Human Development, 45,* 257–263.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S. Y., Pascual, L., et al. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75,* 1115–1139.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32,* 513–531.

Brouwers, S. A., Mishra, R. C., and Van de Vijver, F. J. R. (2006). Schooling and everyday cognitive development among Kharwar children in India: A natural experiment. *International Journal of Behavioral Development, 30,* 559–567.

Brunstein, J. C., Schultheiss, O. C., and Grässmann, R. (1998). Personal goals and emotional well-being: The moderating role of motive dispositions. *Journal of Personality and Social Psychology, 75,* 494–508.

Campbell, D. T. (1986). Science's social system of validity-enhancing collective believe change and the problem of the social sciences. In D. W. Fiske and R. A. Shweder (Eds.), *Metatheory in social science* (pp. 108–135). Chicago: University of Chicago Press.

Carlson, S. M., and Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032–1053.

Case, R., Demetriou, A., Platsidou, M., and Kazi, S. (2001). Integrating concepts and tests of intelligence from the differential and developmental traditions. *Intelligence, 29,* 307–336.

Chao, R. K. (1994). Beyond parental control and authoritarian parenting style: Understanding Chinese parenting through the cultural notion of training. *Child Development, 65,* 1111–1119.

Chao, R. K. (2001). Extending research on the consequences of parenting style for Chinese Americans and European Americans. *Child Development, 72,* 1832–1843.

Chasiotis, A., Bender, M., Kiessling, F., and Hofer, J. (in press). The emergence of the independent self: Autobiographical memory as a mediator of false belief understanding and motive orientation in Cameroonian and German preschoolers. *Journal of Cross-Cultural Psychology*.

Chasiotis, A., Hofer, J., and Campos, D. (2006). When does liking children lead to parenthood? Younger siblings, implicit prosocial power motivation, and explicit love for children predict parenthood across cultures. *Journal of Cultural and Evolutionary Psychology, 4,* 95–123.

AQ12

Chasiotis, A., Keller, H., and Scheffer, D. (2003). Birth order, age at menarche, and intergenerational context continuity: A comparison of female somatic development in West and East Germany. *North American Journal of Psychology, 5,* 153–170.

Chasiotis, A., Kiessling, F., Hofer, J., and Campos, D. (2006). Theory of mind and inhibitory control in three cultures: Conflict inhibition predicts false belief understanding in Germany, Costa Rica, and Cameroon. *International Journal of Behavioral Development, 30,* 192–204.

Chasiotis, A., Kiessling, F., Winter, V., and Hofer, J. (2006). Sensory motor inhibition as a prerequisite for theory of mind: A comparison of clinical and normal preschoolers differing in sensory motor abilities. *International Journal of Behavioral Development, 30,* 178–190.

Chasiotis, A., Scheffer, D., Restemeier, R. and Keller, H. (1998). Intergenerational context discontinuity affects the onset of puberty: A comparison of parent-child dyads in West and East Germany. *Human Nature, 9,* 321–339.

Cheung, F. (2006). A combined emic–etic approach to cross-cultural personality test development: The case of the CPAI. In Q. Jing, H. Zhang, and K. Zhang (Eds.), *Psychological science around the world* (Vol. 2, pp. 91–103). London: Psychology Press.

Church, T. A. (2001). Personality measurement in cross-cultural perspective. *Journal of Personality, 69,* 979–1005.

Cole, K., and Mitchell, P. (2000). Siblings in the development of executive control and a theory of mind. *British Journal of Developmental Psychology, 18,* 279–295.

Cole, M. (1999). Culture in development. In M. H. Bornstein and M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook* (pp. 73–123). Hillsdale, NJ: Erlbaum.

Dana, R. H. (2000). Culture and methodology in personality assessment. In I. Cuellar and F. Paniagua (Eds.), *Handbook of multicultural mental health: Assessment and treatment of diverse groups* (pp. 97–120). San Diego, CA: Academic Press.

Deregowski, J. B., and Serpell, R. (1971). Performance on a sorting task: A cross-cultural experiment. *International Journal of Psychology, 6,* 273–281.

Doi, K. (1982). A two dimensional theory of achievement motivation: Affiliative and non-affiliative. *Japanese Journal of Psychology, 52,* 344–350.

Engelhard, G., Hansche, L., and Rutledge, K. E. (1990). Accuracy of bias review judges in identifying differential item functioning on teacher certification tests. *Applied Measurement in Education, 3,* 347–360.

Fahr, J.-L., Dobbins, G., and Cheng, B. D. (1991) Cultural relativity in action: A comparison of self-ratings made by Chinese and US workers. *Personnel Psychology, 44,* 129–147.

Georgas, J., Berry, J. W., Van de Vijver, F. J. R., Kagitcibasi, C., and Poortinga, Y. H. (Eds.). (2006). *Families across cultures. A 30-nation psychological study*. Cambridge, United Kingdom: Cambridge University Press.

Greenfield, P. M. (1997). Culture as process: Empirical methods for cultural psychology. In J. W. Berry, Y. H. Poortinga, and J. Pandey (Eds.), *Handbook of cross-cultural psychology. Vol. 1: Theory and method* (2nd ed., pp. 301–346). Needle Heights, MA: Allyn and Bacon.

Greenfield, P. M. (2000). Three approaches to the psychology of culture: Where do they come from? Where can they go? *Asian Journal of Social Psychology, 3,* 223–240.

Grimm, S. D., and Church, T. A. (1999). A cross-cultural investigation of response biases in personality measures. *Journal of Research in Personality, 33,* 415–441.

Hofer, J., Busch, H., Chasiotis, A., Kärtner, J., and Campos, D. (2008). Concern for generativity and its relation to implicit power motivation, generative goals, and satisfaction with life: A cross-cultural investigation. *Journal of Personality, 76,* 1–30.

Hofer, J., and Chasiotis, A. (2003). Congruence of life goals and implicit motives as predictors of life satisfaction: Cross-cultural implications of a study of Zambian male adolescents. *Motivation and Emotion, 27,* 251–272.

Hofer, J., and Chasiotis, A. (2004). Methodological considerations of applying a TAT-type picture-story-test in cross-cultural research: A comparison of German and Zambian adolescents. *Journal of Cross-Cultural Psychology, 35,* 224–241.

Hofer, J., Chasiotis, A., and Campos, D. (2006). Congruence between social values and implicit motives: Effects on life satisfaction across three cultures. *European Journal of Personality, 20,* 305–324.

Hofer, J., Chasiotis, A., Friedlmeier, W., Busch, H., and Campos, D. (2005). The measurement of implicit motives in three cultures: Power and affiliation in Cameroon, Costa Rica, and Germany. *Journal of Cross-Cultural Psychology, 36,* 689–716.

Holland, P. W., and Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., Gupta, V., and GLOBE associates. (2004). *Leadership, culture and organizations: The GLOBE study of 62 nations*. Thousand Oaks, CA: Sage.

Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

Inglehart, R. (1997). *Modernization and postmodernization. Cultural, economic, and political change in 43 countries*. Princeton, NJ: Princeton University Press.

Jackson, D. N. (1984). *Manual for the Personality Research Form*. Port Huron, MI: Research Psychologists Press.

Kagan, S., and Knight, G. P. (1981). Social motives among Anglo American and Mexican American children: Experimental and projective measures. *Journal of Research in Personality, 15,* 93–106.

Kagitcibasi, C. (2005). Autonomy and relatedness in cultural context: Implications for self and family. *Journal of Cross-Cultural Psychology, 36,* 403–422.

Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ: Erlbaum.

Keller, H., and Greenfield, P. M. (2000). History and future development in cross-cultural psychology. *Journal of Cross-Cultural Psychology, 31,* 52–62.

Keller, H., Hentschel, E., Yovsi, R. D., Abels, M., Lamm, B., and Haas, V. (2004). The psycho-linguistic embodiment of parental ethnotheories: A new avenue to understand cultural differences in parenting. *Culture and Psychology, 10,* 293–330.

Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., and Papaligoura, Z. (2004). Developmental consequences of early parenting experiences: Self-recognition and self-regulation in three cultural communities. *Child Development, 75,* 1745–1760.

Lillard, A. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123,* 3–32.

Lonner, W. J., and Adamopoulos, J. (1997). Culture as antecedent to behavior. In J. W. Berry, Y. H. Poortinga, and J. Pandey (Eds.), *Handbook of cross-cultural psychology*. *Vol. 1: Theory and method* (2nd ed., pp. 43–83). Needham Heights, MA: Allyn and Bacon.

Marín, G., Gamba, R. J., and Marín, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology, 23,* 498–509.

McClelland, D. C. (1985). *Human motivation*. New York: Cambridge University Press.

McCrae, R. R., Terracciano, A., and 79 Members of the Personality Profiles of Cultures Project. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology, 89,* 407–425.

Miller, J. G. (1997). Theoretical issues in cultural psychology. In J. W. Berry, Y. H. Poortinga, and J. Pandey (Eds.), *Handbook of cross-cultural psychology. Vol. 1: Theory and method* (2nd ed., pp. 85–128). Needham Heights, MA: Allyn and Bacon.

Mullis, I. V. S., Martin, M. O., and Foy, P. (2005). *IEA's TIMSS 2003 International report on achievement in the mathematics cognitive domains. Findings from a developmental project*. Chestnut Hill, MA: Boston College.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22,* 376–398.

Nkaya, H. N., Huteau, M., and Bonnet, J. (1994). Retest effect on cognitive performance on the Raven-38 Matrices in France and in the Congo. *Perceptual and Motor Skills, 78,* 503–510.

Noack, P., Hofer, M., Kracke, B., and Klein-Allermann, E. (1995). Adolescents and their parents facing social change: Families in East and West Germany after unification. In P. Noack, M. Hofer, and J. Youniss (Eds.), *Psychosocial responses to social change* (pp. 129–148). Berlin: Walter de Gruyter.

Öngel, U., and Smith, P. B. (1994). Who are we and where are we going? JCCP approaches its 100th issue. *Journal of Cross-Cultural Psychology, 25,* 25–53.

Organisation for Economic Development and Cooperation. (2003). *The PISA 2003 assessment framework—mathematics, reading, science and problem solving knowledge and skills*. Paris: Organisation for Economic Development and Cooperation.

Phalet, K., and Lens, W. (1995). Achievement motivation and group loyalty among Turkish and Belgian youngsters. In M. L. Maehr and P. R. Pintrich (Eds.), *Advances in motivation and achievement: Vol. 9. Culture, motivation, and achievement* (pp. 31–72). Greenwich, CT: JAI Press.

Poortinga, Y. H. (1989). Equivalence of cross-cultural data: An overview of basic issues. *International Journal of Psychology, 24,* 737–756.

Poortinga, Y. H., and Van de Vijver, F. J. R. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology, 18,* 259–282.

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral Brain Sciences, 1,* 515–526.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rogoff, B. (2003). *The cultural nature of human development.* New York: Oxford University Press.

Rohner, R. P. (1984). Toward a conception of culture for cross-cultural psychology. *Journal of Cross-Cultural Psychology, 15,* 111–138.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., and Lee, K. (2006). The development of executive functioning and theory of mind. A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17,* 74–81.

Scheier, I. H. (1959). The method of natural variation in applied educational research. *Journal of Educational Research, 52,* 167–170.

Schliemann, A., Carraher, D., and Ceci, S. J. (1997). Everyday cognition. In J. W. Berry, P. R. Dasen, and T. S. Saraswathi (Eds.), *Handbook of cross-cultural psychology*: *Vol. 2. Basic processes and human development* (2nd ed., pp. 177–216). Needham Heights, MA: Allyn and Bacon.

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

Scribner, S., and Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.

Segall, M. H. (1984). More than we need to know about culture, but are afraid to ask. *Journal of Cross-Cultural Psychology, 15,* 153–162.

Segall, M. H., Campbell, D. T., and Herskovits, M. J. (1966). *The influence of culture on visual perception*. Indianapolis, IN: Bobs-Merrill.

Shebani, M. F. A., Van de Vijver, F. J. R., and Poortinga, Y. H. (2005). A cross-cultural test of the phonological loop hypothesis. *Memory and Cognition, 33,* 196–202.

Shen, C., and Pedulla, J. J. (2000). The relationship between students' achievement and their self-perception of competence and rigour of mathematics and science: a cross-national analysis. *Assessment in Education, 7,* 237–253.

Shikanai, K. (1978). Effects of self-esteem on attributions of success-failure. *Japanese Journal of Experimental Social Psychology, 18,* 47–55.

Smith, P. B., Harb, C., Lonner, W., and Van de Vijver, F. J. R. (2001). JCCP between 1993 and 2000: Looking back and looking ahead. *Journal of Cross-Cultural Psychology, 32,* 9–17.

Smith, P. B., Peterson, M. F., and Schwartz, S. H. (2002). Cultural values, sources of guidance and their relevance to managerial behaviour: A 47 nation study. *Journal of Cross-Cultural Psychology, 33,* 188–208.

AQ13 Stanat, P., and Luedtke, O. (in press). Multilevel issues in international large-scale assessment studies on student performance. In F. J. R. van de Vijver, D. A. van Hemert, and Y. H. Poortinga (Eds.), *Individuals and cultures in multilevel analysis*. Mahwah, NJ: Erlbaum.

Steinberg, L., Lamborn, S. D., Dornbusch, S. M., and Darling, N. (1992). Impact of parenting practices on adolescent achievement: Authoritative parenting, school involvement, and encouragement to succeed. *Child Development, 63,* 1266–1281.

Super, C. M. (1983). Cultural variation in the meaning and uses of children's "intelligence". In J. B. Deregowski, S. Dziurawiec, and R. C. Annis (Eds.), *Expiscations in cross-cultural psychology* (pp. 199–212). Lisse, The Netherlands: Swets and Zeitlinger.

Super, C. M., and Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *International Journal of Behavioral Development, 9,* 545–569.
Takata, T. (1987). Self-depreciative tendencies in self evaluation through social comparison. *Japanese Journal of Experimental Social Psychology, 27,* 27–36.
Tanzer, N. K., Gittler, G., and Ellis, B. B. (1995). Cross-cultural validation of item complexity in a LLTM-calibrated spatial ability test. *European Journal of Psychological Assessment, 11,* 170–183.
Triandis, H. C. (1976). Methodological problems of comparative research. *International Journal of Psychology, 11,* 155–159.
Triandis, H. C. (1995). *Individualism and collectivism*. Boulder, CO: Westview Press.
Tulviste, P. (1991). *The cultural-historical development of verbal thinking.* Commack, NY: Nova Science.
Van de Vijver, F. J. R. (2000). The nature of bias. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 87–106). Mahwah, NJ: Erlbaum.
Van de Vijver, F. J. R., Daal, M., and Van Zonneveld, R. (1986). The trainability of abstract reasoning: A cross-cultural comparison. *International Journal of Psychology, 21,* 589–615.
Van de Vijver, F. J. R., and Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.
Van de Vijver, F. J. R., and Leung, K. (2001). Personality in cultural context: Methodological issues. *Journal of Personality, 69,* 1007–1031.
Van de Vijver, F. J. R., and Poortinga, Y. H. (2002). On the study of culture in developmental science. *Human Development, 45,* 246–256.
Van de Vijver, F. J. R., and Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology, 54,* 119–135.
Van Hemert, D. A., Van de Vijver, F. J. R., Poortinga, Y. H., and Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences, 33,* 1229–1249.
Van Langen, A., Bosker, R., and Dekkers, H. (2006). Exploring cross-national differences in gender gaps in education. *Educational Research and Evaluation, 12,* 155–177.
Whiting, B. (1976). The problem of the packaged variable. In K. Riegel and J. Meacham (Eds.), *The developing individual in a changing world* (Vol. 1, pp. 303–309). The Hague, The Netherlands: Mouton.
Winter, D. G. (1996). *Personality: Analysis and interpretation of lives*. New York: McGraw-Hill.
Yu, A. (1996). Ultimate life concerns, self, and Chinese achievement motivation. In M. H. Bond (Ed.), *The handbook of Chinese psychology* (pp. 227–246). Hong Kong: Oxford University Press.
Zhang, J., and Bond, M. H. (1998). Personality and filial piety among college students in two Chinese societies: The added value of indigenous constructs. *Journal of Cross-Cultural Psychology, 29,* 402–417.

Author Queries

AQ1. "Bornstein and Lansford, Chapter 14, this volume" correct citation here, or are you referring to Chapter 27?
AQ2. "We explain below..." okay as changed to "We explain later..." because the term "below" may not apply in pages?
AQ3. Sentence beginning "In line with such arguments..." okay as edited?
AQ4. Sentence beginning "Various statistical techniques are available..." okay as edited?
AQ5. Did you intend a question mark as part of this header? If so, should it perhaps read "How Do We Approach Culture?'' or "How Should Culture Be Approached?''
AQ6. Sentence beginning "Cultur and psyche OK as edited?"
AQ7. I found a publication year of 2007 for the Keller book Cultures of Infancy. Okay to replace "in press" with "2007" here, in other citations of the book, and in the reference list?
AQ8. In the section under the header ,"Culture and data analysis'', the one xtremely long paragraph that comprised the section was broken into 3 to improve readibility. OK?
AQ9. Sentence beginning "Studies using a molecular approach..." okay as edited?
AQ10. TAT okay as written out as Thematic Apperception Test?
AQ11. Sentence beginning "A few models..." OK as edited?
AQ12. Please update Chasiotis, A., Bender, M., Kiessling, F., and Hofer, J. (in press) reference if possible.
AQ13. Please update Chasiotis, A., Bender, M., Kiessling, F., and Hofer, J. (in press) reference if possible.