

HANDBOOK OF RESEARCH ON MULTICULTURAL EDUCATION

SECOND EDITION

James A. Banks

Editor

Cherry A. McGee Banks

Associate Editor

2004



JOSSEY-BASS

A Wiley Imprint

www.josseybass.com

QUANTITATIVE METHODS IN MULTICULTURAL EDUCATION RESEARCH

Amado M. Padilla

Stanford University

Quantitative educational research with ethnic minorities has a long history. The earliest studies with educational implications focused on the intellectual assessment and school achievement of African American, immigrant, and other ethnic minority students (Kamin, 1974; Valencia & Suzuki, 2001). This research legacy is now well known for its failure to consider many variables that are critical for assessing student ability (Valencia & Suzuki, 2001; Padilla, 1988). For example, in assessing intelligence, IQ tests were given special status, and it was assumed that instruments such as the Stanford Binet Test could be used to uncover differences in intellectual ability between individuals or racial groups. With the special status ascribed to IQ tests, little attention was given to the fact that in the development of IQ tests minority children were not included in standardizing the instrument (Kamin, 1974). Further, in actually carrying out research on differences between groups on IQ tests, little if any attention was given to social class or language background of the subject, or cultural differences between the groups being compared. Researchers today recognize the many problems inherent with the older body of IQ-related research, but great reliance is still placed on standardized tests of achievement; many of the same problems still exist in properly using and interpreting findings from these tests.

More important is the recognition of a set of assumptions inherent in the older IQ studies that are still operative today in educational research involving ethnic minorities. Complex, interrelated, and conforming to commonsense qualities that make them appealing, these identifiable assumptions are that (a) the White middle-class

American is the standard against which other groups should be compared; (b) the instruments used for assessing differences are universally applicable across groups, with perhaps only minimal adjustment for culturally diverse populations; and (c) although we need to recognize such sources of potential variance as social class, educational attainment, gender, cultural orientation, and proficiency in English, these are nuisances that can later be discarded.

This chapter challenges these assumptions and offers numerous suggestions for improving quantitative research with ethnic minority respondents. There are many forms of quantitative research: controlled experiment, quasi-experiment, survey, observational study, case study, statistical simulation, meta-analysis (a study of studies), and others. Some of these are hypothesis-generating, meaning that they explore data to form or sharpen a hypothesis about a population so as to assess future hypotheses. Some are hypothesis-testing, which means they seek to assess specific a priori hypotheses or estimate parameters by random sampling from a population.

In this chapter, the primary focus is on nonexperimental quantitative research since this is the primary mode of research in education (Johnson, 2001). However, the methodological difficulties to be identified in conducting research with ethnic populations are also applicable to experimental studies involving a treatment and control group. In addition, problems of instrumentation and measurement of constructs are discussed in a way that is applicable to research in general, regardless of the specific methods used.

THE SOCIAL CONSTRUCTION OF KNOWLEDGE

Before these issues can be investigated meaningfully, it is important to examine how the construction of knowledge has proceeded in the social sciences and what impact this has had on the study of minority populations generally, as well as in education specifically. Within this discussion, questions and challenges are directed at Eurocentric paradigms that have dominated our approach to accumulating scientific facts. Critique of Eurocentric approaches to the study of ethnic minority populations in social science research, and in educational research particularly, is based on the fact that such approaches have resulted in erroneous interpretations because of specific biases inherent in the paradigms themselves.

In a discussion of quantitative research with minority populations, it is important to address the topic of scientific paradigms used by researchers to define their approach. A starting point for our discussion is the distinction between universalistic and relativistic methods. Allport (1937) framed the central tenets of these two methods with respect to their use in psychology, but the distinction is equally appropriate to educational research since educational researchers employ the same paradigms.

Universalistic Versus Relativistic Approaches to Educational Research

The universalistic approach seeks confirmation of general truths that extend across cultural groups. In contrast, the relativistic approach seeks to uncover a particular truth that is confined to a single culture or social group. Since its initial formulation, the debate has been recast in terms of etic and emic principles (Matsumoto, 1994). Proponents of the universalistic or etic view believe that concepts and methodologies are basically valid across different cultures. Conversely, the relativistic or emic view maintains that concepts and methodologies do not have universal validity; they may be appropriate only within a narrow range of cultural groups.

Educational researchers have generally followed the social sciences in their adoption of acceptable paradigms that rest on a universal framework (Banks, 1993; Kerlinger, 1979). According to the universalistic framework, theory and hypothesis testing should guide research. Thus, quantitative methods are employed and statistical inferences are used to draw conclusions that support the universal principles. In addition, advocates of a universalistic framework also adopted a research strategy that calls for the logic of the laboratory, where experimentation, control, and random assignment of subjects to experimental and control groups is possible (Dehue, 2001). It wasn't long before it was apparent to social scientists and educational researchers alike that a strictly

experimental research approach was not possible in the real world of people, schools, and large-scale societal problems such as class size reduction or curriculum reform. However, Campbell and Stanley (1966) showed that quasi-experiments could be designed in a natural context, with varying degrees of experimental or statistical control and where the necessary statistical assumption of random assignment of respondents to groups could be maintained. Dehue presents an invaluable historical critique of the idea of randomized controlled studies.

The important concern here is that this approach eschews the importance of such variables as culture; rather, emphasis is on a comparative approach that uses similar measures to compare males and females, children of different age or ethnic groups, etc. Investigators who use this approach argue that universal principles can be uncovered only by means of a comparative approach to research.

The universal approach has come under sharp criticism from numerous sources, notably feminist and minority researchers, because of its Eurocentric perspective (for a summary of this critique, see Banks, 1993). The most salient feature of the Eurocentric paradigm is its focus on a monocultural, male-oriented, and comparative approach to research (Yoder & Kahn, 1993). White male researchers from a monocultural perspective using White and generally middle-class students as the normative population have developed the majority of instruments and research procedures used in educational research. These instruments and procedures are then used primarily to assess some psychological or educationally relevant construct with a White (male and female) middle-class student population. This approach lends itself to a narrow database, resulting in biased conclusions of substantive educational outcomes that are problematic even for White samples that differ from the normative population. The problem is worse if use of the instruments and procedures is extended to ethnic minority populations who do not share all the demographic characteristics of the normative group (e.g., immigrant students who are not proficient in English). A related point has been made by Sears (1986), who showed how research based on college students tested in academic laboratories on academic-like tasks has culminated in social psychological theories that are incompatible with the everyday life experiences of most non-college-age majority-group adults.

There is nothing wrong with this approach to research so long as whatever groups are being compared are equivalent in all demographic characteristics including social class, cultural background, and proficiency in English. However, there is still room for caution when, for example, men and women from the same social class and cultural background are compared on a task requiring interpersonal competition and men are found to score

higher on competition (Griffen-Pierson, 1990). Is this difference due to the use of an inappropriate male-oriented task, or a real difference between the sexes resulting from a genetic disposition found in males but not females? It should be obvious that an informed person would agree with the former and not the latter conclusion. Yet in comparisons of cultural or ethnic groups, conclusions similar to the latter "genetic" interpretation can be found, with no consideration for the fact that the instruments may have been biased to begin with because of a conceptual framework and set of tasks that favored one group over the other (Griffen-Pierson).

The problem arises when "biased" instruments that favor non-Hispanic whites are used in a comparative research framework to examine differences between racial or ethnic groups (Azibo, 1996; Rogler, 1999; Sue, 1999). The comparative research framework requires a statistical test between at least two groups that have been equated on all variables known to have an influence on the behavior in question (Plutchik, 1974). However, if both the construct being assessed and the method for assessing the construct originate in the same cultural context, then a comparative approach may seriously increase the potential for research bias. Thus it would be inappropriate to use the instrument developed for one cultural group to assess group differences, as necessitated by the comparative research framework (Sue, 1999). Proof that such comparisons are the mainstay of much of the empirical research in education is evident from a cursory examination of our major professional journals in education.

The situation is even more problematic because the comparative research framework assumes that there is some standard by which comparisons are made. Although not always stated explicitly, the standard is usually the non-Hispanic White middle class, and any deviations are interpreted negatively as deficits or differences that possibly require intervention. In sum, no fewer than three potentially harmful consequences can be identified when a White, privileged male norm is adopted: (a) overgeneralization of findings that benefit members of the "norm" group, (b) exaggeration of differences that extend beyond the true nature of the between-group difference, and (c) evaluation of deficiency levied at a low-status contrast group.

In sum, studies driven by a universalistic approach have the potential to blind the researcher to important between-group differences that go uncontrolled, the result of which is erroneous conclusions if mean score differences are found to be statistically significant. Many minority students are subjected to institutionalized practices of assimilation (English-only instruction, academic tracking, and so on) that may have the effect of devaluing self-esteem and academic achievement. Consequently, between-group differences in academic performance as measured by

grade point average or standardized test scores (SAT and others) might be a reflection of unmeasured variance (such as unequal educational opportunity) that is due to extraneous factors not taken into account by researchers (Lee, 2002).

A Paradigm Shift in Ethnic Research

African American researchers (Azibo, 1996; Parham, White, & Ajamu, 1999; and others) have challenged the Eurocentric approach because it typically assumes a White standard reference group. This has led to an emic or Afrocentric approach employing conceptual categories and worldviews adopted from traditional African cultures that serve as the standard for understanding African Americans. According to Azibo (1996), the Afrocentric approach challenges comparative studies that employ theory or methods designed to maintain African American inferiority.

In a less philosophical and more methodological challenge to the universalistic or etic approach, Marín and Marín (1991) and Rogler (1999) argue that researchers must be knowledgeable of Hispanic culture and demographic information to carry out useful research with this population. Yee (1992) makes a similar case in discussing the stereotypes and misperceptions that persist when Asian Americans are studied by educational researchers.

Rogler (1989), for example, argues in favor of a culturally sensitive research approach that places culture at the center of the research enterprise. According to Rogler:

Research is made culturally sensitive through a continuing and open-ended series of substantive and methodological insertions and adaptations designed to mesh the process of inquiry with the cultural characteristics of the group being studied. . . . The insertions and adaptations span the entire research process, from the pretesting and planning of the study, to the collection of data and translation of instruments, to the instrumentation of measures, and to the analysis and interpretation of the data. Research, therefore, is made culturally sensitive through an incessant, basic, and active preoccupation with the culture of the group being studied throughout the process of research. (p. 296)

It is important to understand that the merits of the scientific method are not being challenged here. What is being called into question is the claim to objectivity by some researchers who believe that their scientific paradigms are neutral as far as minority groups (including women) are concerned. The scientific method consists of a series of paradigms, each governed by distinct assumptions, rules, and methods of conducting research (Kuhn, 1970). However, a culturally sensitive approach to research holds that empirically derived facts are not valid for all time but need to be examined from the perspective of the assumptions, language, and activities of the community of scientists. If

the "community" has not included minority researchers, then it is small wonder that culture and other salient characteristics of ethnic groups are not considered important in mainstream research. A more emic-oriented approach thus has the responsibility to examine these "established facts" through a new lens.

In sum, if we begin with Rogler's (1989) definition of culturally sensitive research and couple this with the criticisms of the Eurocentric paradigm discussed above, a new paradigm is called for, in which the study of a specific ethnic group is valued for its own sake and that group need not be compared to another. Proponents of this view of research argue that it is valuable first to research and understand within-ethnic-group differences; and only when intragroup heterogeneity is well understood is comparative ethnic-group research meaningful. An alternative way of stating this is that if majority group researchers are not admonished for including *only* a White sample in their study, why should ethnic researchers be cautioned against not including a White comparison group in their research?

We shall now turn to discussion of major challenges in conducting research with ethnic minority groups. Important considerations that are required of the researcher intending to conduct quantitative research with ethnic groups are identified and assessed in the next sections of this chapter.

CHALLENGES IN CONDUCTING RESEARCH WITH ETHNIC GROUPS

There are numerous problems in conducting research with ethnic minority populations, many of which are frequently overlooked by investigators unfamiliar with research topics that may be particularly sensitive for these populations. A useful guide for conducting research in ethnic minority communities is available from the American Psychological Association (2000). Some of these issues are related to identifying and selecting a sample. Because of the diversity of the U.S. population, the researcher has to be careful not to confound culture, ethnicity, and social class in selecting a sample. A related concern is the failure to recognize heterogeneity (for instance, differences in level of acculturation) existing within an ethnic group, which can lead to variation in outcome measures and result in misinterpretation of the findings. Finally, we consider language and culture barriers confronting researchers in ethnic minority communities.

Properly Identifying, Describing, and Selecting a Sample

When minority students are included in quantitative research, it usually results—intentionally or not—in

documenting the low academic achievement of Hispanic, African American, and Native American students in comparison to non-Hispanic white students (Lee, 2002) and the higher mathematics attainment of Chinese, Korean, and Japanese students compared to Caucasian students (Flynn, 1991; Sue & Okazaki, 1990). Many educational studies involving Hispanic and African American students examine them from the perspective of their failure in the educational system, or how to improve our understanding of factors associated with (under)achievement, such as achievement motivation (Marchant, 1991) and parenting styles (Dornbusch, Ritter, Leiderman, Roberts, & Fraleigh, 1987; Lamborn, Mounts, Steinberg, & Dornbusch, 1991). There are, however, a few excellent quantitative studies of the educational achievement of language-minority, usually Spanish-speaking, students in bilingual education programs (August & Hakuta, 1997).

Few studies have examined ethnic minority students with respect to their success in education. It is rare to find research such as Kraft's (1991) study of what makes a Black student successful on a predominantly White campus; Alatorre Alva's (1991) examination of the academic invulnerability of Mexican American high school students; Arrelano and Padilla's (1996) study of highly successful undergraduates at Stanford University and the role played by personal, family, and school resources in understanding high academic attainment of Mexican American students; Duran and Weffer's (1992) study of the influential family and school factors associated with the achievement of successful Mexican American immigrant high school students; or Strom, Johnson, Strom, and Strom's (1992) investigation of programs for gifted Hispanic children and their parents. Specialized topics such as desegregation and bilingual education also include ethnic minority students, but empirical research in these areas is scant.

Understanding the characteristics of the population is critical if we are to generalize the results properly and replicate the findings. However, many studies do not describe the subject population sufficiently to enable replication. Two examples of subject description from research journals are presented here to illustrate the lack of information that would enable true replication:

1. "Subjects were 32 children from the Berkeley area. Subjects included 15 boys and 17 girls. Children came from a variety of socioeconomic and ethnic backgrounds, though most were Caucasian and middle class" (Gopnik & Meltzoff, 1992, p. 1094).
2. "In all, 423 sixth- and seventh-grade students . . . participated in the study. The school was in a predominantly working-class [geographic location here in original] community. The average student age was 11.87 and 13.08 for sixth and seventh graders,

respectively. The sample was equally representative of males (52%) and females (48%), with 68% of the sample being Caucasian, 23% Black, 5% Hispanic, and 7% other minority status (Wentzel, 1991, p. 1068).

Each description offers the basic information about number of students, gender, age, social status, and ethnicity. However, the information is much too general, and each variable is described for the sample as a whole. To be specific, it is not known whether the social class of the non-Hispanic white students was similar to that of the other ethnic groups. What does predominantly, mostly White or Caucasian, and middle-class mean? Does it mean 51% of the sample, or something closer to 100%? Do the categories of White and Caucasian include Hispanic, as they should? Does a sample of 60% White and 40% Black from diverse social backgrounds mean 50% White middle-class, 35% Black working-class, and the remaining White and Black subjects representing other social classes? Are the Hispanic students from English- or Spanish-speaking homes? Are the Hispanic students of Mexican, Cuban, Puerto Rican, or Guatemalan heritage? Do they belong to some other Latino national group? Are they biracial? Does diverse social background include children from homeless families or poverty conditions? What about single-parent versus two-parent families?

In the latest edition of the *APA Publication Manual* (American Psychological Association, 2001), authors are given quite explicit guidelines about the detail required when describing subjects in a study:

Report major demographic characteristics such as sex, age, and race/ethnicity, and, where possible and appropriate, characteristics such as socioeconomic status, disability status, and sexual orientation. When a particular demographic characteristic is an experimental variable or is important for the interpretation of results, describe the group specifically—for example, in terms of national origin, level of education, health status, and language preference (if applicable). . . . Even when a characteristic is not an analytic variable, reporting it may give readers a more complete understanding of the sample (pp. 18–19).

Clearly, this level of detail cannot always be attained. However, if researchers know beforehand that they are going to employ a diverse subject pool in their research, they can make provision to collect demographic information that can be used to enhance their description of the population studied. This has two benefits. First, by gathering important demographic information the researcher is in a better position to interpret the findings in terms of context of culture, social class, acculturation level, and length of residence in the United States. Second, by adequately describing the subject population, other interested researchers can replicate the study using a comparable population. The advantage of replication is that it

allows researchers to test the robustness of findings and thereby adds an element of truthfulness to them if replication results in similar outcomes. However, it is commonplace to find little information regarding culturally diverse respondents in the methods section of an article. Apparently many authors, peer reviewers, and journal editors do not question the absence of critical information when it comes to a culturally diverse population.

In identifying and selecting a sample, keep in mind that several major methodological issues are essential:

- What important demographic characteristics of the population does the researcher need to know in order to properly interpret findings and make generalizations?
- Are the procedures for sample selection adequately described so that replication can be carried out?
- If the sample was not randomly selected and is therefore not representative of the population, what limitations does this pose for generalizing?

One critical issue in identifying a population is to understand its demographic characteristics. What diversity is represented? Understanding the heterogeneity that exists in various communities is essential if the researcher is to understand how best to move forward in gaining information about culturally different respondents. Further, if we are to lessen the potential for bias found in research with ethnic minority individuals (and found in what passes as truth in the field), we need to rethink our designation of respondents. Bond (1988) discusses the bias of labels and perspectives in research with culturally diverse populations. Designations such as “deprived background,” “disadvantaged,” “lack of stimulation,” “poverty,” and the more current “limited English proficient” and “at-risk learners” have been widely associated with culturally diverse individuals. As Bond states:

Discussions of the effects of poverty on development frequently equate minority membership with the poor. Although it is important to acknowledge the disproportionate representation of certain ethnic and racial groups at the lower socioeconomic levels and the significance of such environments to these groups’ development, this oversimplified equation of minority status with poverty perpetuates stereotypes and obscures the factors that contribute to this relationship. (p. 46)

The frequent failure of researchers to understand that poverty imposes constraints on the choices a person or family has available also causes problems. Poverty limits the educational, social, and recreational choices a family can make for their children, so it is no small wonder that poorer children do less well academically than middle-class children. The fact that low SES results in limited educational choices is frequently overlooked when between-ethnic-group comparisons are made.

When studies do incorporate culturally diverse individuals, it is generally within a cross-racial/ethnic comparison. In many such comparisons, ethnicity is confounded with social class, whereby the comparison is made between a middle-class White population and a working-class African American or Hispanic (perhaps immigrant) population. Such glaring differences between groups would not be acceptable in any other area of education-related research. Imagine a study of differences on the SAT equating girls from upper-class backgrounds attending a private school with boys from working-class families enrolled in a public school vocational track. A reviewer of such a study would immediately recognize the problem of trying to infer gender differences from a sample that ignores differences in socioeconomic status between the two groups of students and would likely recommend that a peer-refereed journal reject the paper. Yet as Sue (1999) points out, this level of critique on the part of a journal reviewer may not occur in a research study involving ethnic groups.

In addition, as mentioned previously, the subject section of an article might not describe the background of the research subjects sufficiently so that one knows what "Hispanic" or "Asian" means. Is the subject a third-generation American of Chinese ancestry from a middle-class home, or an immigrant from a working-class mainland Chinese family who arrived illegally in the United States? These two individuals differ in fundamentally important ways, and to gloss over these differences by merely indicating that the sample consisted of Asian Americans is to miss the important point being made here.

In order that culture, ethnicity, and social class not be confounded, it is important to understand the unique cultural features of the group. For example, length of residence in the United States is important for immigrants; language usage and preference is also significant if for no other reason than to determine what language to use in data collection. Another variable is the ethnic or racial self-identification of the person. If a respondent prefers *Pilipino* to Filipino or Pacific Islander, then this conveys important information about political orientation and level of acculturation to U.S. culture and customs. When studying children, parents' level of education (or human capital) is often critical in understanding how and why children perform as they do on standardized tests. Children from homes with more human capital perform at a higher level than children whose parents possess less education.

Few researchers have actually taken these factors into account when conducting research. Laosa (1982) demonstrated, in a study of Anglo and Chicana mothers' teaching strategies, that any differences in teaching strategies between the two ethnic groups disappeared when the mother's level of formal education was a controlling factor. Similarly, Gutierrez and her colleagues (Gutierrez &

Sameroff, 1990; Gutierrez, Sameroff, & Karrer, 1988) examined the heterogeneity of acculturation and social class in a study of Mexican American and Anglo mothers' concepts of development. They included Anglo upper-class and lower-class groups and six Mexican American groups representing various socioeconomic status and low, medium, and high acculturation. Results showed significant SES effects for both Anglo and Mexican American mothers, and considerable within-group variability for the Mexican American group. In addition, comparisons between the Mexican American highly acculturated and Anglo American, both higher SES, indicated that Mexican American mothers gave more cognitively complex responses.

In another study of low-income Mexican mothers, Richman, Miller, and LeVine (1993) also demonstrated that "maternal schooling emerges from this study as an important influence on maternal responsiveness during infancy in and of itself, rather than as reflecting the social variables with which it is often associated" (p. 62). The significance of these research studies is in demonstrating the varying effects of both educational background and acculturation level in responses of Mexican American subjects, and even between similar Mexican American subgroups (grouped by acculturation level).

Several researchers have included ethnic minority individuals in their studies so that the subject sample would be representative of a particular geographic area. This representative sampling approach can be useful in understanding some component of behavior or development as reflected by the diversity of individuals in the community. An example of a representative sample is Stevenson, Chen, and Uttal's (1990) study of the achievement of 3,000 first-, third-, and fifth-grade Black, White, and Hispanic children, and a subsample of the parents' beliefs about academic achievement. They selected "20 elementary schools covering the range of socioeconomic and ethnic groups within the area. Two classrooms each at first, third, and fifth grades within each school were randomly selected for study" (p. 509). Subsamples of students were selected for individually administered tests. Stevenson et al. provided information about the age of students by ethnic group and the percentage of students in each group born in the United States. In addition, they presented information on family structure, language spoken at home, educational level, and family income. Black, Hispanic, and White families differed significantly in both education and occupation level, though multiple comparison analyses were not employed to examine specific differences among the groups. The researchers analyzed student achievement according to both ethnicity and SES (mother's level of education). Results indicated that ethnic differences in children's mathematics performance, but not reading performance, were no longer significant when mother's

education level was controlled. There were many ethnic differences in parental attitudes, but unfortunately, effects resulting from educational differences were not analyzed. Thus the confounding effects of education were examined in children's academic performance, but not in parental beliefs, despite the previously identified significant differences in education level among the ethnic groups.

This example illustrates one of the problems in conducting research with ethnic populations. The issue of confounding is substantial even in a carefully conducted study. Even when results include statistical controls, findings are not always discussed with respect to the confounding. Thus the results may be representative of the population, but the comparisons made among groups may be inappropriate because of the serious problem of confounding.

We have just seen some of the problems that may arise in representative samples. What about random samples? Culver, Wolffe, and Cross (1990) wanted to obtain a random sample from a population of 9,753 teachers who were identified as early in their career (those with six years of full-time teaching or less) and who were currently teaching. However, a simple random sample would not be likely to yield enough African American teachers for the researchers' purposes because teachers from this racial group made up only 13.5% of the population. Thus they decided to "sample at random approximately equal numbers of Blacks and Whites (actually 350 Whites and 375 Blacks)" (p. 329). Since it was not possible, in this case, simply to select a random sample from the population, another sampling procedure was used that included a random sample of White teachers and an oversample of African American teachers. In many research projects, especially of the survey type, investigators should seriously consider the advantages of oversampling of an ethnic group to have a large enough sample size to carry out statistical tests of significance.

In another work, Finn and Achilles (1990) conducted an experimental study of class size and used random sampling effectively in composing their classrooms:

All school systems in the State of Tennessee were invited to participate. . . . About one third of the districts, representing 180 schools, expressed an interest in participating. After negotiation, the final sample consisted of 76 elementary schools that were large enough to have at least three kindergarten classes. . . . Within each school, children entering kindergarten were assigned at random, by the project staff, to one of three class types: small, with an enrollment range of 13-17 pupils; regular, with an enrollment of 22-25 pupils; or regular with aide, with 22-25 pupils but with a teacher aide formally assigned to work with the class. Teachers were assigned at random to classes as a separate step. (pp. 559-560)

Classes were categorized by composition, as containing White students, all minority students, or a mixture of

White and minority. They were also classified by location, as inner-city, urban, suburban, or rural. However, within this well-controlled experimental study was a serious confound concerning race and ethnicity, location, and social class (participation in the free lunch program was considered as designating low income). There were no inner-city classes with all White students and no suburban or rural classrooms with only minority students. Only five classrooms were inner-city and mixed. There was also a "strong association between minority status and participation in the free lunch program. About 70% of the student sample is either minorities receiving free lunches, or Whites not receiving free lunches; Yule's Q association measure is .78" (Finn & Achilles, 1990, p. 561). Consequently, in this "randomized experiment" there is certainly helpful information about class-size effects, but the effects for ethnic minority status are seriously confounded with location and social class. Fortunately, the researchers recognized this problem and conducted their analyses to take this confound into consideration.

In sum, researchers need to take seriously the problem of confounding variables when they do research with ethnic minority populations. How a sample is selected for study can greatly influence the results and hence the generalizations that can be made on the basis of the findings. Moreover, if sufficient attention is given beforehand to possible confounding variables while selecting a subject population, misinterpretation of findings is reduced (Wilkinson, 1999). When an ethnic population is studied, there is often a confound between social class and cultural group membership. To state the obvious, a difference in reading achievement between African American students and non-Hispanic White students does not mean that African American students come from a culture that does not value learning to read, but it could mean that because of poverty the home conditions are not optimal for early reading development.

Understanding the Heterogeneity Within an Ethnic Group

Similar to the problem of confounding is the lack of understanding by most researchers of the heterogeneity within an ethnic minority population. Some researchers state that they include Asians without going into further detail about who these Asians are (Yee, 1992). Are they middle-class, or working-class? Are they Chinese, Japanese, Hmong, Korean, Vietnamese, or Cambodian? Were they born in the United States? If not, how long have they lived in this country? Similarly, for Hispanics, there are considerable differences among Mexican American, Cuban, Puerto Rican, Argentinean, Chilean, Colombian, Guatemalan, and Central American Mayans (for whom Spanish is a second language). Within each ethnic subgroup are social-class (e.g., educational background),

acculturation, and language differences (particularly dialect and language variations among Central American Indians).

To illustrate the point made in the preceding paragraph, consider how Keefe and Padilla (1987) showed that newly arrived Mexican immigrants differed in many ways from second- and third-generation Mexican Americans. Puerto Ricans who live on the mainland differ from those who commute between the island and the mainland, as well as from those who have always lived on the mainland (Rodriguez, 1989). The Cubans who fled Castro in the 1960s differ in social class and color (White) from the poor and largely Black Cuban exiles who came in the 1970s (Suarez, 1993). There are Black Hispanics who are frequently identified by the majority group as African American, without any understanding of their Latino roots. Berndt, Cheung, Lau, Hau, and Lew (1993) showed that perceptions of parenting differed among Chinese parents living in mainland China, Taiwan, and Hong Kong. In the United States, where Chinese (and Asians in general) are frequently treated as though they are homogeneous, this study demonstrates important socialization differences among Chinese families on the basis of country of origin. Also of importance is the growing number of interracial children. Increasingly in our schools, there are children who are of mixed ethnic or racial heritage. It is important for researchers to be aware of mixed-heritage individuals in conducting research because these individuals often have loyalty to both of their heritage backgrounds and cultures (Yee, 1992). These examples all illustrate complex intragroup differences that must be understood to facilitate replicability and generalizability.

Difficulties from Cultural and Language Barriers

Another difficulty in conducting research with ethnic minority populations concerns language and cultural barriers. Superficial speculation on this point suggests that one can always get someone to translate instruments and broker with community members, but issues are more intricate than the simple term *language and cultural barriers* suggests. As asserted in the previous section, there are a number of subject-selection and subject-description issues that can introduce serious methodological flaws into research, though these serious flaws do not often draw the attention of journal peer reviewers. How can an outsider interpret the results of a study involving ethnic respondents if individuals who understand the ethnic community are not included in a significant capacity on the research team (Rogler, 1999)?

De la Luz Reyes and Halcón (1988) make this assertion regarding research on Hispanics:

As Hispanic academics, our research interests often stem from . . . a compelling need to lend a dimension of authenticity to the prevailing theories about our communities. Said another way, we want to provide our own perspectives regarding prevailing negative assumptions about our values, culture, and language. . . . Our interest in these research areas is also motivated by a concern for assisting our community in improving its second-rate status in the education, economic, and political arenas. Tired of reading about ourselves in the social science literature written by non-minorities, we want to speak for ourselves, to define, label, describe, and interpret our own condition from the "inside out." We feel strongly about providing a balance to the existing literature and research on Chicanos. (p. 306)

Marín and Marín (1991) make a similar point in their book *Research with Hispanic Populations*, noting that "some [Hispanic] community members perceive social science research as a form of exploitation in which non-minority individuals reap the benefits of the data collection effort" (p. 42).

In addition, Marín and Marín (1991) delineate several other cultural barriers that make it difficult for minority and nonminority researchers alike to work within an ethnic minority community. Although their comments are aimed at Hispanics, many of these points are relevant for other ethnic communities as well:

Suspicion of government involvement in a research project is more likely when individuals or their family members and friends have lived in political climates where oppressive governments make use of informers and home visits to gather compromising information to be used in surveillance, social control, or other abuses of a person's rights. In addition, many Hispanics, regardless of their immigration status (documented, undocumented, refugee, parolee), live in fear of being stopped by agents of the Immigration and Naturalization Service and of being asked to document their citizenship or immigration status. . . . Also of relevance in determining the rate of cooperation with an investigation is the type of personal or community benefit that is to be accrued by participating in the study. (pp. 43-44)

In recent years, with the backlash against newcomers observed through nativist movements such as "English only"; with state restrictions designed to limit social services to immigrants; and with new federal laws calling for greater restrictions in immigration policies, especially after the September 11, 2001, attacks on the World Trade Center and the Pentagon, immigrants are more cautious than ever about volunteering to participate in any type of research project.

Language barriers can also be problematic for researchers. For example, Spanish is not the same in all Hispanic communities. Hispanics from numerous countries may be able to communicate with one another, but there are significant dialect differences that should be reflected in letters of introduction, human subject

consent forms, questionnaires, and other written and oral forms of communication with the group of interest. Issues of instrument translation are discussed in the next section of this chapter.

Some cultural and language barriers can be surmounted by including members of the community on the research team. However, including a Mexican American graduate student raised in a middle-class suburban neighborhood who learned Spanish in high school in a project of Latino immigrant parents and children may not constitute inclusion of an informed community member in the research enterprise. Similarly, including a Cuban or Puerto Rican educated on the East Coast as an interviewer may not be so helpful with a Mexican American population in South Texas, particularly with respect to language issues, even if the interviewer is perfectly fluent in Spanish; the reverse would be equally inappropriate. As another example, a highly acculturated Filipino American may have difficulty collecting data from recently arrived Filipinos on attitudes toward schooling. Unless investigators are culturally aware of traditional customs, they could ask questions in such a way as to promote inaccurate responses or to offend the respondent. Thus to minimize culture and language barriers it is necessary to have thorough cultural knowledge of the target population.

It is equally important to know something about the language background and proficiency level of individuals who serve as translators of materials or interviewers for non-English speaking respondents. I am reminded of a case where an investigator working with Chinese immigrants in the San Francisco area assumed that all Chinese spoke the same dialect. Consequently, when he showed up with a Mandarin-speaking interviewer to meet with parents who spoke Cantonese, he was both embarrassed by his lack of familiarity with the very community he wished to study and disappointed that his effort to identify and train an interviewer had been wasted.

In sum, the challenges in researching ethnic minority communities are considerable. I do not want to leave the impression that only minority researchers should study minority communities. The take-home message is that unless the researcher knows the community well, it is critical to include members of the community in the research study as true partners and not just as translators, interviewers, or data coders.

Community leaders are often more vested in the research conducted in their neighborhood or school than the actual researcher. But they want to see the research being used in ways that benefit the community, not just ways that benefit the researcher's career or the interests of the sponsoring agency. For this reason, it makes good sense to identify community advocates and to find ways to make them legitimate partners in the research process.

INSTRUMENTATION AND MEASUREMENT

Critical to any quantitative study are the issues of instrumentation and measurement. Regardless of whether one uses a rating scale, an inventory, a standardized achievement test, or any other type of performance-based outcome measure, the issues are the same. That is, the instrument must be appropriate for measuring change resulting from an educational program or intervention. If the research involves a survey questionnaire, then the items must reliably assess how the sample population evaluates the items. Accordingly, this section includes a general discussion of these instrumentation concerns because of their relevance in identifying suitable measures when conducting research with an ethnic population.

One issue in research with ethnic groups concerns identifying *appropriate* outcome measures. Many instruments may be suitable for White middle-class subjects but not for a culturally diverse sample. In identifying and selecting outcome measures, one has to consider the psychometric qualities of the instrument. There are several questions that should provoke serious consideration whenever we use an instrument with ethnic respondents:

- Are the selected instruments reliable and valid with the ethnic group in question? Is there equivalence of meaning of key concepts across cultures (e.g., self-esteem, independence-dependence, etc.) used in educational research?
- Do we have to translate the instrument into another language? If so, what is the best way to do this?
- Is it necessary to use specially designed instruments to assess such characteristics as acculturation, ethnic identity, English-language proficiency, or culture-specific learning strategies? How are such instruments identified for use with an ethnic population?
- Do ethnic respondents answer paper-and-pencil instruments such as a questionnaire or personal inventory in the same manner as majority-group respondents? Are there response biases that researchers need to be knowledgeable of?

Are the Instruments Appropriate?

Many studies have examined instruments to assess their suitability for a particular ethnic minority population. For example, Knight, Tein, Shell, and Roosa (1992) evaluated the cross-ethnic equivalence of parenting and family interaction measures between Hispanic and non-Hispanic White families. In the study they assessed four instruments: the Children's Report of Parental Behavior Inventory, the Parent-Adolescent Communication Scale, the Family Adaptability and Cohesion Evaluation Scale II,

and the Family Routines Inventory. These instruments were selected because the authors believed them to be most useful in large-scale field assessments that would include ethnic minorities. They examined the interaction measures by using small panels of Hispanic individuals with some training in measurement. The examination involved having panel members evaluate each item for cultural relevance, that is, "the degree to which the behaviors and attitudes reflected in the items were applicable in the Hispanic culture" (p. 1394). Panel members also evaluated the items according to their underlying construct using two formats. Then the panel members "identified three rejection items, three cohesion items, three adaptability items, and seven family routine items as potentially irrelevant or as questionably relevant for the Mexican American culture" (p. 1394). As Knight et al. (1992) point out, "The explanations provided by the panel members of the lack of relevance for each item fell into one of two categories: (1) the item itself, or some wording or phrasing in the item, either has an ethnically specific meaning or has unclear meaning to members of the Mexican American culture; and (2) the item was worded poorly or vague terms were used, such that it is unlikely that subjects would understand the meaning of the item regardless of their ethnicity" (p. 1394).

The conclusion from this panel approach was that "there appears to be a small subset of items . . . that are likely of limited item equivalence because the behaviors or attitudes represented in these items are of limited applicability or generalizability to the Hispanic family" (Knight et al., 1992, p. 1395). The study then went on to assess the item equivalence and functional equivalence of the latent structure and subscale intercorrelations between Hispanic and non-Hispanic samples. From the findings, it was clear that some scales, minus certain subscales, had sufficient cross-ethnic equivalence for English-speaking Hispanic samples, while other scales required further scale development.

Many other scales have been assessed for their appropriateness for various cultural groups. Two scale categories are used to illustrate the psychometric work that has been conducted to address the issues of instrument appropriateness: (a) achievement scales, and (b) self-esteem scales. Numerous other types of scale could have been selected for discussion here, but achievement and self-esteem scales have a long history of use with ethnically diverse populations and serve to exemplify some points that we wish to emphasize regarding proper use of instruments with ethnic respondents.

Achievement Scales. Considerable research has examined achievement differences between ethnic minority and nonethnic minority students (Valencia & Suzuki, 2001). However, it is often not clear whether the various

achievement measures in use are comparable in what they measure and in how they should be used in the research. Frisby (2001) has produced a readable account of the many types of achievement test currently available to researchers interested in culturally diverse students. For example, there are individual standardized achievement tests used to evaluate individual students who are believed to be in need of some form of special treatment in school because of a suspected learning disorder, or for determining giftedness. There are group standardized achievement tests such as the National Assessment of Educational Progress (NAEP) test, the California Test of Basic Skills, the Stanford Achievement Test, etc. These are standardized tests used broadly across the country to monitor trends in achievement by grade level, content area (reading, math), gender, ethnic differences, and social class of the learner.

Achievement measures have been translated into other languages, but rarely are the instruments carefully assessed for their comparability. For example, the Metropolitan Achievement Test (MAT) was translated into Chinese and is used by Chinese bilingual programs because there are no alternative measures. Similarly, the Comprehensive Tests of Basic Skills (CTBS) have a Spanish version, the Spanish Assessment of Basic Education (SABE). The SABE was developed by Spanish-speaking experts and normed with a native Spanish-speaking U.S. student population. However, the SABE is not equivalent to the CTBS. The SABE was not developed as a translation of the CTBS, but as a separate measure that was to be comparable to the CTBS; careful studies have not been conducted as to how comparable the two versions are. There are also other Spanish-language achievement tests: *La Prueba Riverside de Realización en Español (La Prueba)* and *Aprenda: La Prueba de Logros en Español*. Both are norm-referenced with Spanish-speaking populations in the United States.

Aside from the comparability across languages, there is the issue of content comparability for middle-class English-speaking non-Hispanic white students and for culturally and linguistically diverse students. In an important conceptual article, Helms (1992) has argued that cognitive ability test items are not all culturally equivalent for African Americans and Whites. As she points out, there is an "absence of clearly articulated, theoretically based models for examining the influence of race-related cultural factors on cognitive ability" (p. 1089). Helms maintains that cultural equivalence of standardized cognitive ability testing remains to be studied. The problem is partially because no commonly accepted alternatives to statistical approaches for instigating such an investigation are available. More problematic still is the inability of psychologists and psychometricians to articulate the relevant issues as they affect test takers from various racial and ethnic groups within the United States. Therefore, the conclusion that

whatever construct is measured by standardized CATs (Cognitive Ability Tests) constitutes universal intelligence or general cognitive ability for all racial and ethnic groups in this country is dubious at best.

Achievement tests suffer from the same validity and cultural bias problems discussed earlier. Even criterion-referenced achievement tests based on students' classroom curriculum fall short. In Stevenson et al.'s (1990) study of the achievement of African American, White, and Hispanic children, the researchers carefully constructed criterion-based achievement measures of mathematics and reading. The results showed significant ethnic differences in reading, but not mathematics, even after carefully controlling for mother's education background. Stevenson et al.'s conclusion clearly identifies the problem associated with achievement tests and even criterion-referenced tests:

Our interpretation of this finding [of significant ethnic differences in reading achievement after controlling for mother's level of education] is that the content of the material the children were asked to read was based on experiences and knowledge that were less likely to be part of the daily lives of the Black and Hispanic than of the White children. Comprehending the meaning of text is difficult when the topics lie outside the child's everyday experience. Hispanic children bore the additional burden of being asked to read a language that typically was not the native language of their parents. The content of reading classes, more than that of mathematics classes, reflects situations that exist in the dominant culture. Minority children may be penalized in reading because the materials require information to which they have had less exposure outside of school than the White children have had. However, our use of tests based on the textbooks to which the children had been exposed may have been responsible for reducing the magnitude of ethnic differences in this study compared to studies that have relied on standardized tests of achievement. Typically, standardized achievement tests are based on what children are expected to know, rather than on what they necessarily have encountered. (p. 520)

In sum, the point made by Stevenson et al. (1990) is that important contextual, experiential, and cultural factors need to be considered when assessing achievement across ethnic and cultural groups. Much progress along these lines has already taken place from the time when little to no effort was made to ensure that the content of a test was equivalent for all groups who took it. The practical significance of this has taken on greater importance today, with increased implementation of high-stakes testing in schools for purposes of teacher and student accountability such as teacher merit pay and high school exiting.

Self-Esteem Scales. Another category of instrument that has been used extensively in educational research with minority students encompasses the many self-esteem scales that are available. There are some important issues exemplified by self-esteem measurement that affect psychosocial development in children and

adolescents. One concerns the definition of self-esteem. Theorists have long debated how to conceptualize and measure self-esteem (Rosenberg, 1979). Two major conceptual camps are distinguished, in terms of whether we view self-esteem as a unidimensional construct of self-worth (global view) or whether we evaluate an individual along several domains in addition to an overall global self-worth, or differentiated, view (for a review, see Harter, 1999).

These theoretical perspectives have not incorporated considerations relevant to culturally diverse students. Early literature suggested that ethnic minority children showed a lower level of self-esteem than White majority group children (for reviews, see Rosenberg, 1979; Wylie, 1979), a finding contested on several methodological and conceptual grounds (Wylie, 1979). However, despite the frequency with which self-esteem measures are used, there is still little theory that is relevant to ethnic minority students. As Clark (1965) suggested four decades ago, living in a racist and oppressive society that keeps people of color in poverty impinges on how they appraise their self-worth. For example, to continue with Harter's (1999) formulation of self-esteem, adolescents describe themselves in terms of characteristics from many domains: academic competence, athletic competence, job competence, physical appearance, social acceptance, close friendships, romantic appeal, and so on. Self-appraisal on many of these domains is made in the context of social comparison (I am smart because I am taking honors U.S. history and calculus, but I am not an athlete because I couldn't make the varsity basketball team!). As a result, depending on an adolescent's academic, social, and athletic standing in school, self-appraisal of worth may vary widely among all these domains (Harter, 1999). Further, minority adolescents in a school where there are few students from their same ethnic community may appraise themselves even more negatively on a larger number of domains because of their "outsider status" and possibly because of perceived prejudice and discrimination directed at them by peers.

Martinez and Dukes (1987) have conceptualized self-esteem according to the differentiated viewpoint, which suggests that ethnic minority students may evaluate themselves differentially among two major domains. They hypothesized that ethnic minority students would evaluate themselves lower than majority group members on public aspects of self-esteem (such as intelligence) but rate themselves high in private domains (satisfaction with self, for one). The rationale for this hypothesis was that in the public domain of self-esteem the majority group is the standard, whereas in the private domain the individual or the ethnic group is the standard. In their study, Martinez and Dukes found support for their thesis. The private-domain self-esteem ratings of African American and Hispanic students were higher than those of non-Hispanic

White students, while in the public domain of intelligence, which is measured in terms of the majority school-based culture, Hispanic and African American students rate themselves lower than non-Hispanic White students.

Scales for measuring self-esteem among ethnic minority students also fall short. In addition to the typical psychometric problems, they do not assess any cultural items that may affect self-esteem, and they are oriented toward middle-class norms. For example, in Harter's (1982) Perceived Competence and Social Acceptance Scale for Young Children (Grades 1-2), there is an item that asks children whether they have spent the night at a friend's house. In determining a child's social competence (which Harter views as a part of self-esteem), those first- and second-grade children who have stayed overnight at their friends' homes get higher scores than those who have not and thus attain a higher social competence score. Clearly, this item may be biased against some children whose culture is less permissive about sleepovers. Many ethnic children, especially girls, are not allowed to spend the night away from home unless it is the home of a trusted family member, and certainly not the home of a schoolmate whom the family does not know intimately. Children from traditional homes (Chinese, Muslim, Mexican, Vietnamese) may have stayed at their grandparents' or other family members' homes overnight, but this would not be credited on the Harter scale.

A study of Chinese children in Taiwan using the Perceived Competence Scale for Children (a translated version) reported the factorial validity of the scale for the Chinese sample (Stigler, Smith, & Mao, 1985). As with White American samples, there was a high correlation between the perceived cognitive competence and actual achievement. However, in keeping with Chinese culture and concerns of modesty, children tended to underrate their competence compared to White American children. In addition, unlike White American children, Chinese children differentiated satisfaction with self from the desire to change for the better. Stigler et al. conclude, "Whereas idealized perceptions of the self might reflect social desirability bias among American children, this same bias might produce self-effacement among Chinese children" (p. 1269). A similar finding would not be unlikely with Chinese American and other Asian background children living in the United States.

In a related study, Rotenberg and Cranwell (1989) assessed the self-concept of Native American and White American children using the "20 statements" test, an open self-description measure. They found that Native American children referred more frequently in their open description to kinship roles, traditional customs and beliefs, and moral worth than did White American children.

Thus, the findings differ with the instrument in comparison of ethnic minority and White children. As a

whole, these results clearly point out that theories of self-esteem need to take into account how culture influences the manner in which a person defines his or her role in the social group (Dana, 1997). Also important is how the level of acculturation interacts with the social context to influence the self-esteem of culturally diverse individuals. Only when scales have been developed that incorporate culture into the items will we really understand how children evaluate themselves. Similar findings would result from an examination of other categories of scales.

Ethnic-Specific Instruments

The second set of issues concerns the question of whether it is necessary to use a specially designed instrument to assess characteristics such as acculturation, ethnic identity, or stress. As we have seen in the preceding discussion, mainstream scales that assess cognitive ability, achievement, and psychosocial development have not been generated with ethnic minority populations in mind. They have largely been produced by and for a majority-group middle-class population. The problem becomes even more complex when we focus on culturally specific behaviors or areas of development (such as ethnic identity, acculturation, or acculturative stress) that have not been viewed as significant issues for most White middle-class individuals.

As we have discussed, it is important for purposes of research to recognize the heterogeneity within certain ethnic groups. One important source of heterogeneity that has received considerable attention in recent years is acculturation (Moyerman & Forman, 1992; Rogler, Cortes, & Malgady, 1991; Shen & Takeuchi, 2001). It is well known that with contact between majority-group members and immigrants, the newcomers and their offspring eventually acquire the language, values, beliefs, and behaviors of the majority group. Theorists such as Berry (1990); Bourhis, Moise, Perreault, and Senecal (1997); Olmedo (1979); and Padilla (1980) have discussed the conceptual and methodological issues involved in the study of acculturation. Many questions have arisen regarding the process of acculturation and such considerations as gender, age of immigration, educational level, length of residence, and extent of intergroup contact with members of the majority group. More recently, Landrine and Klonoff (1996) have shown that African Americans also demonstrate differences in level of acculturation that are unrelated to social class and educational attainment.

There is no agreed-upon universal scale for measuring acculturation; neither is there, for that matter, any agreed-upon best scale for use with a particular ethnic group. However, most scales can be characterized by two general item categories: (a) self-rated proficiency and use of the home language, and (b) preference for ethnic-related

activities and friends. By way of illustration, numerous scales can be found in the literature for use with diverse ethnic groups: African Americans (Landrine & Klonoff, 1996), Koreans (Kim, 1988), Mexican Americans (Cuelar, Harris, & Jasso, 1980), Asian Americans (Suinn, Richard-Figueroa, Lew, & Vigil, 1987), and American Indians (Oetting & Beauvais, 1990, 1991).

The important consideration regarding acculturation, for our purposes, is that the relationship among culture change, psychosocial adjustment, and educational attainment is in need of more attention. Most of the research on acculturation involves various indexes of mental health such as depression (Moyerman & Forman, 1992) and, more recently, the relationship between acculturation and ethnic identity (Buriel & Cardoza, 1993; Marín, 1993).

There is evidence in the literature suggesting that immigrant and second-generation students who are more traditionally oriented perform better academically than their later-generation and more acculturated counterparts (Buriel & Cardoza, 1988; Caplan, Whitmore, & Choy, 1989; Padilla & Gonzalez, 2001; Portes & Rumbaut, 2001; Suárez-Orozco, 1989). Thus immigrant youths who identify with their ethnic group and who are more traditional in beliefs and values have better grades and are more likely to go on to college than their acculturated peers. For example, Portes and Rumbaut (2001) found that for Southeast Asian students high grade point average was related to how their parents answered four questions on an acculturation measure. The questions pertained to preservation of culture and identity; "sticking together" for social support and mutual assistance; living where there are people of their own ethnic group; and no interest in returning to their country of origin. Using a 6-point scale from "strongly disagree" (0) to "strongly agree" (6), Rumbaut found that high parental scores were positively related to student high school GPA.

A similar finding appears to hold true for Mexican immigrants (Buriel, 1984; Buriel & Cardoza, 1988; Padilla & Gonzalez, 2001), Punjabi Sikh students (Gibson, 1988), and Central American children (Suárez-Orozco, 1989). A complicating factor also has to do with whether we are describing a situation of accommodation without assimilation, or a form of biculturalism in which students acquire English-language proficiency, know the culture of their parents and of the school, and have friends from different ethnic groups. As a consequence, they feel more comfortable in school and at home and do better overall on their academic work (Alatorre Alva, 1991).

In sum, the finding that immigrant students are more motivated to study and have more positive school attitudes than later-generation ethnic students is important. Whether, however, how school achievement is influenced by ethnic acculturation via a strategy of accommodation without assimilation or biculturalism is still an open

question. Much more research is required before we fully understand the relationship for immigrant students between school performance and acculturation.

Another consideration in ethnic-related research has to do with the attitude that minority-group members have toward their own ethnicity. Phinney (1990) and Bernal and Knight (1993) have given us useful reviews of the relevant literature on ethnic identification. As Phinney states, there is much research on how majority group members stereotype minorities, but much less on how minority-group members perceive themselves. The issues here have to do with the evaluation of self-worth in a social context that frequently discriminates against or disparages ethnic groups.

The difficulty with ethnic identification research to date is that there are widely differing approaches to the study of ethnic identity, since groups vary in their experiences as members of a minority group. This has resulted in diverse measurement instruments designed to assess ethnic identity. For example, Phinney and Rotheram (1987) and Bernal and Knight (1993) offer a useful starting point for understanding the various avenues that have been pursued in ethnic identification research.

Assessing the Response Patterns of Ethnic Respondents

The third set of questions in this section has to do with how ethnic respondents answer questions on objective instruments that do not coincide with the intent or content of the instrument. This is a serious matter that merits extensive discussion. One aspect of this question in the research literature concerns what is described as *response set preferences* in answering questions on various types of objective instrument. For example, some respondents choose the extremes on Likert-type scales, while others prefer the middle choice. Other informants respond to questions even if they have no opinion on the topic addressed in the survey instrument. Still other subjects offer *socially desirable or acquiescent* answers (yea-sayers) on an instrument or during an interview. These respondents use strategies in answering that make them look good in the eyes of the examiner or that reduce the possibility of more questions. Whatever the reason for response bias in answering, the crucial matter is that such bias results in error that can be consequential because it raises questions about the validity of the data obtained from respondents.

Response Set. Bachman and O'Malley (1984) have shown that African Americans have a preference for selecting the extreme responses on instruments that use a Likert-type scale. According to Bachman and O'Malley, this is why African Americans have sometimes been found to be higher in self-esteem than Whites. This

extreme response set means that African Americans use the extreme scores (both positive and negative) more than Whites. The result, then, is a distribution of scores on a measure of self-esteem that may be a reflection of the response set rather than of actual between group differences on self-esteem.

Hui and Triandis (1989) have found that Hispanics are also more likely to use an extreme response set on a 5-point Likert scale than are non-Hispanic Whites. Similar findings are discussed by Marín and Marín (1991), who state that such extreme responding is particularly evident with low-aculturated Hispanic respondents. It is unknown whether other ethnic groups follow a pattern similar to African Americans and Hispanics in answering Likert-type questions.

The reason for being concerned about whether extreme response set has occurred is that such responding can seriously affect the results and interpretation of a study by giving a misleading impression of group variances. According to Bachman and O'Malley (1984), one solution to the problem of extreme set responding is to collapse the extreme category on each end of the scale (collapse "disagree" with "strongly disagree" on one end of the scale, and "agree" with "strongly agree" on the other). Scores between ethnic and majority group subjects then become more similar. However, before collapsing the extreme categories of a 5-point scale the investigator needs to be aware that compressing a scale from five to three categories results in what is no longer an interval scale, even though we may use it as such. A more satisfactory way to address the question of biased response pattern is to ensure that any scale includes a healthy dose of reverse-keyed items.

Another question that arises in this context that has remained largely unexplored is whether the race of the examiner influences response bias on the part of ethnic minority respondents. In other words, do minority respondents offer more or less face-saving positive responding when the examiner is of a race other than the respondent's? Thus, in the absence of a theory to explain why ethnic respondents prefer the extreme response categories, it may simply be sufficient to examine the pattern of responses on a Likert scale to determine whether extreme responding has occurred and to note this in reporting the results.

Social Desirability and Acquiescence. Another concern in conducting research with ethnic respondents has to do with the possible problem of social desirability or acquiescence in responding. Social desirability refers to the tendency to "deny socially undesirable traits and to claim socially desirable ones" (Nederhof, 1985, p. 264); responding in a socially desirable manner may occur consciously or unconsciously (Paulhus, 1984). Acquiescence

refers to a type of responding wherein respondents agree (yea-saying) with statements presented to them regardless of their content. Whether or not social desirability or acquiescence responding is deliberate matters little, since either way it creates a major concern in assessing the validity of self-reported measures (Rogler, Mroczek, Fellows, & Loftus, 2001).

The question of social desirability is raised here because ethnic differences in the tendency to offer socially desirable responses have been reported in the literature. For instance, Ross and Mirowsky (1984) administered a questionnaire with a battery of measures—including social desirability, locus of control, and psychological distress—to a sample of non-Hispanic Whites and Mexican Americans in El Paso, Texas. An additional sample of Mexicans from Juarez, Mexico, was administered the same battery of instruments, but in Spanish. It was found that the greatest level of social desirability was reported by the Mexican sample, followed in turn by the Mexican Americans, and finally the non-Hispanic Whites. Ross and Mirowsky also found an inverse relationship between social class and social desirability; individuals lower in socioeconomic status were most likely to present a pattern of socially desired responses.

Ross and Mirowsky (1984) interpreted their findings by suggesting that as we move down the socioeconomic ladder, acquiescence appears as a self-presentation strategy of those who are relatively powerless in society. According to these authors, people of greater powerlessness attempt to present a good face to those members of society whom they perceive to be higher in social standing in an effort to be more accepted in society.

In a reanalysis of four data sets that included responses by nearly 2,000 Hispanics and more than 14,000 non-Hispanic Whites, Marín, Gamba, and Marín (1992) reported that Hispanics showed a greater tendency to agree with items than did Whites. Two variables were found by Marín et al. to correlate with acquiescence responding. The first was educational level; those respondents, regardless of ethnicity, who possessed fewer than 12 years of formal schooling showed more response acquiescence than did the more highly educated respondents. The other variable was acculturation; it was found that Hispanics who were more acculturated evidenced less response acquiescence. An important cultural interpretation is offered by Marín and Marín (1991) to explain the findings. According to them, Hispanic culture promotes social acquiescence through the social script of *simpatia*, which "mandates politeness and respect and discourages criticism, confrontation, and assertiveness. Providing socially desirable answers could be a way to promote positive, smooth relationships between researcher and participant" (p. 106).

These few studies demonstrate the importance of understanding how the cultural background that the ethnic

respondent brings to the task of completing an interview, survey, or questionnaire of various types determines the response patterns that emerge. Equipped with this understanding, the investigator might anticipate responses quite different from those obtained from the White respondents on whom most instruments are standardized. Clearly, more research is required on the question of ethnic differences in response patterns on objective measurement instruments.

Another consideration in this discussion pertains to approaching use of an instrument with members of an ethnic group for which the scale was not normed. It is always a good practice to determine the adequacy of such a scale with the ethnic group in question. We recommend at least two methods for doing this: Cronbach alpha (internal-consistency reliability) and exploratory factor analysis. In the earlier discussion of a study by Knight et al. (1992), we showed how Knight and his colleagues tested their instruments for their appropriateness with ethnic samples. The discussion that follows elaborates on the approach taken by Knight et al. and offers suggestions for using instruments appropriately.

Internal-Consistency Reliability. It is good practice for researchers to question the reliability of their instruments whenever they conduct a study involving ethnic samples. At a minimum, Cronbach alpha for internal-consistency reliability should be computed on the scales used in a study. This should be done for each of the ethnic groups *separately* if two or more groups are being compared. On the basis of the resulting alpha coefficients, the researcher must decide whether to proceed with the study or search for a more appropriate instrument.

A frequently asked question is, What is the minimum acceptable level of reliability to gauge the suitability of a scale for use with a sample? Pedhazur and Pedhazur-Schmelkin (1991) maintain that the acceptability of a reliability estimate depends on the "decisions made on the basis of the scores and the possible consequences of the decisions" (p. 109). Thus, the reliability of an instrument should be as high as possible (minimum of $r = .70$) for more consequential (high-stakes) program or policy decisions, but it can be lower ($r = .50$) for research purposes involving low-stakes outcomes to be made about assessing differences between groups on a psychological or educational measure where no intervention is planned.

If the reliability estimates are low for an ethnic group, then it is always the responsibility of the investigator to point out that the estimates were low and offer caution regarding any interpretations to be drawn from the study. An item-by-item analysis may also enable the researcher to understand why a particular scale is more tenuous for one population than for another.

In conducting quantitative research with ethnic populations, it is also essential to determine, whenever possible,

whether the constructs being measured by the instruments have the same meaning for each ethnic group being studied. Depending on the instrument in question and the sample size, it may be possible to examine the construct validity of instruments across groups by means of exploratory factor analysis.

Exploratory Factor Analysis. It is commonplace today for many of our educational and psychological scales to be developed using methods of factor analysis. For example, a researcher may have a theory about the underlying construct of learning anxiety, or about which attitudinal predispositions are important in learning a foreign language (Gardner, 1985). Armed with a theory, the researcher develops a set of items that appear to measure the constructs of learning anxiety and attitude toward learning a second language. The items are then formatted into an objective questionnaire and (usually) arranged using a Likert scoring continuum; respondents are instructed to check whether they agree with each item on the continuum, which is arranged from "strongly agree" to "strongly disagree." The instrument is administered to a large number of respondents and the data analyzed by means of a factor analysis. Factor analysis is a data-analytic procedure for arriving "at a relatively small number of components that will extract most of the variance of a relatively large set of indicators (variables, items)" (Pedhazur & Pedhazur-Schmelkin, 1991, p. 598).

There are several types of factor analysis, but the one most commonly used in the literature is called principal components factor analysis with Varimax rotation. The important thing about this statistical procedure is that it produces clusters of items that are statistically independent of each other. The researcher is then able to examine the items within a cluster and determine whether they fit the construct the researcher has in mind. For example, if the researcher believes that two separate constructs (say, anxiety and predisposition toward second-language learning) underlie the learning of English by nonnative speakers, she can then examine the clusters of items to determine whether the two key constructs are reflected in the items. Technically speaking, those items that contribute to (or load on) a component (or factor) using some criterion level such as 0.30 are then retained and those that do not are discarded. In this way, scales are developed that can be refined further to determine whether they measure the construct in question.

To continue our example, suppose we have developed an instrument using factor analysis that has two subscales: one measures learning anxiety and the other attitudinal disposition toward second-language acquisition. We then hypothesize that a good second-language learner would be characterized as showing little learning anxiety and a positive attitude toward learning a new language.

We can test this hypothesis by administering our new instrument to a large group of Latino immigrant students enrolled in high school ESL classes. As dependent measures, we could use class grades and a measure of English language oral proficiency. If we find that students who score low on anxiety and high on disposition to learn English also have higher grades and higher oral proficiency assessments in English than students who show a pattern of high anxiety and low disposition to learn English, then we may have confidence in our instruments and in their utility to measure our key constructs of anxiety and attitude toward learning of a second language.

The proof of how good our instrument is, however, depends on whether the constructs hold up in the same way with other groups of English language learners. Thus we might then administer our questionnaire to a group of newcomer Pakistani students. However, bear in mind that the original items were most likely developed with Spanish speakers in mind and written and administered in Spanish. Now, to use the instrument with a new group of English language learners, the items must be translated into Urdu (following the guidelines for translation discussed earlier in this chapter). After the translated questionnaire has been administered to the students, a new factor analysis should be carried out to ensure that our underlying constructs of anxiety and predisposition to learning a second language are applicable to this new population of English language learners. Accordingly, if factor analysis indicates a similar factor structure as that found with Spanish speakers, then we can be confident that our measuring instrument is suitable for cross-cultural generalization. Bear in mind that the factor loadings (the items) may not be identical in our two respondent groups. However, if there is sufficient similarity in how the items load on the two subscales, then we can have confidence in our constructs, in how we have measured them, and in the translation equivalence across two quite distinct languages (Pedhazur & Pedhazur-Schmelkin, 1991).

On the other hand, if a factor structure emerges that is different from that found with the Latino students, we need to rethink our constructs and how we define them operationally so that they are measurable. Perhaps there is an issue in the way we have attempted to arrive at equivalence of meaning in our translation of the instrument. This is a common problem in cross-cultural research. Or perhaps the worldviews and attitudes toward learning generally of the other sample populations are so distinct that our constructs do not make sense to one of the groups of students for whom we are attempting to find the best pedagogical strategies for teaching English.

In sum, quantitative research with culturally and linguistically diverse populations is not always easy. Issues of instruments and their external validity and reliability

with diverse populations must be a core concern to the educational researcher intent on studying ethnic groups (Sue, 1999). The issues are by no means insurmountable, and various strategies have been offered for thinking constructively about the appropriateness of instruments and ways of collecting data that will prove to be meaningful to the investigator and the community of respondents involved in the research (Messick, 1995).

CONCLUSION

This chapter has covered a number of critical issues that must be considered in doing quantitative research with ethnic minority populations. The chapter opened with discussion of the social construction of knowledge and how two research approaches have shaped, in different ways, how quantitative research is conducted. Central to this discussion is the Eurocentric paradigm. In recent years, it has been called into question because of potential bias in favor of White middle-class (male) college students, who are used as the standard by which to evaluate research findings and to draw inferences to a broader population. Critics of this approach (Sears, 1986; Sue, 1999) have pointed to the dangers for both social sciences generally and cross-cultural research specifically. In opposition to this approach, ethnic minority researchers and educational scholars have called for a shift away from the Eurocentric paradigm and moved toward more ethnic-sensitive paradigms (Afrocentric, for one). Advocates of these new paradigms maintain that standards should lie with the specific ethnic group in question and that they should be based on the values and worldviews of that ethnic group alone. In addition, researchers (Azibo, 1996; Marín and Marín, 1991; Rogler, 1989) have called for more culturally sensitive approaches to quantitative research with ethnic communities.

In line with the culturally sensitive approaches to quantitative research, the critical challenges to conducting quantitative research with ethnic groups were identified and discussed in the chapter. These challenges involve (a) the importance of identifying, describing, and selecting a sample; (b) understanding of the heterogeneity within an ethnic group; and (c) the difficulties posed by language differences. The importance of each challenge was discussed, pointing out that they are not insurmountable and in fact pose no serious threat to the integrity of a research study so long as there is understanding of the ethnic group being studied.

A recommendation was made that, whenever possible, members of the ethnic community under study be incorporated into the planning and implementation of the research project. This increases the potential for more relevant research questions and approaches. Further, a more

appropriate or bias-free sample may be a likely outcome if the ethnic community is involved in the research enterprise. Also, interpretation of certain findings can be enhanced if "community insiders" are part of the research team.

The final section of this chapter was devoted to issues of instrumentation and measurement. Quantitative research is only as good as the data on which it is based, and this means special attention must be given to the instruments used in research with ethnic populations. The importance of measurement was illustrated by a review of research involving achievement and self-esteem scales. This was followed by a discussion of acculturation and ethnic identification, which have emerged as two central constructs in ethnic-related research. It was shown that there are conceptual reasons for giving significant attention to acculturation and ethnic identification in

our research. This section closed with a discussion of response bias and patterns of socially desirable (acquiescence) responding that have been found in some research with African American and Hispanic informants. This could be a serious concern both in interpreting findings and in deciding what strategies should be followed in future development of instruments for use with ethnic populations.

The chapter closed with two analytic strategies to determine whether the ethnic informants show similar patterns of reliability and interpretation of specific items on a scale. These strategies involved internal-consistency reliability and exploratory factor analysis. This discussion was intended to suggest that the research base with ethnic minority populations can be improved by giving serious attention to the psychometric properties of all of the instruments used in educational research.

References

- Alatorre Alva, S. (1991). Academic invulnerability among Mexican-American students: The importance of protective resources and appraisals. *Hispanic Journal of Behavioral Sciences*, 13, 18-34.
- Allport, G. (1937). *Personality: A psychological interpretation*. New York: Henry Holt.
- American Psychological Association. (2000). *Guidelines for research in ethnic minority communities*. Washington, DC: Author.
- American Psychological Association. (2001). *Publication Manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- Arellano, A. R., & Padilla, A. M. (1996). Academic invulnerability among a select group of Latino university students. *Hispanic Journal of Behavioral Sciences*, 18, 485-507.
- August, D., & Hakuta, K. (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.
- Azibo, D. (1996). *African psychology in historical perspective and related commentary*. Trenton, NJ: African World Press.
- Bachman, J. G., & O'Malley, P. M. (1984). Black-White differences in self-esteem: Are they affected by response styles? *American Journal of Sociology*, 90, 624-639.
- Banks, J. A. (1993). The canon debate, knowledge construction, and multicultural education. *Educational Researcher*, 22(6), 4-14.
- Bernal, M. E., & Knight, G. P. (Eds.). (1993). *Ethnic identity: Formation and transmission among Hispanics and other minorities*. Albany, NY: State University of New York Press.
- Berndt, T. J., Cheung, P. C., Lau, S., Hau, K., & Lew, W.-J. F. (1993). Perceptions of parenting in mainland China, Taiwan, and Hong Kong: Sex differences and societal differences. *Developmental Psychology*, 29, 156-164.
- Berry, J. (1990). Psychology of acculturation: Understanding individuals moving between cultures. In R. Brislin (Ed.), *Applied cross-cultural psychology* (pp. 232-253). Thousand Oaks, CA: Sage.
- Bond, L. A. (1988). Teaching developmental psychology. In P. A. Bronstein & K. Quina (Eds.), *Teaching a psychology of people: Resources for gender and sociocultural awareness* (pp. 45-52). Washington, DC: American Psychological Association.
- Bourhis, R. Y., Moise, L. C., Perreault, S., & Senecal, S. (1997). Towards an interactive acculturation model: A social psychological approach. *International Journal of Psychology*, 32, 369-386.
- Buriel, R. (1984). Integration with traditional Mexican American culture and sociocultural adjustment. In J. L. Martinez, Jr., & R. H. Mendoza (Eds.), *Chicano psychology* (2nd ed., pp. 95-130). Orlando, FL: Academic Press.
- Buriel, R., & Cardoza, D. (1988). Sociocultural correlates of achievement among three generations of Mexican American high school seniors. *American Educational Research Journal*, 25, 177-192.
- Buriel, R., & Cardoza, D. (1993). Mexican American ethnic labeling: An intrafamilial and intergenerational analysis. In M. E. Bernal & G. P. Knight (Eds.), *Ethnic identity: Formation and transmission among Hispanics and other minorities* (pp. 197-210). Albany, NY: State University of New York Press.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.
- Caplan, N., Whitmore, J. K., & Choy, M. H. (1989). *The boat people and achievement in America*. Ann Arbor, MI: University of Michigan Press.
- Clark, K. B. (1965). *Dark ghetto: Dilemmas of social power*. New York: HarperCollins.
- Cuellar, I., Harris, L. C., & Jasso, R. (1980). An acculturation scale for Mexican American normal and clinical populations. *Hispanic Journal of Behavioral Sciences*, 2, 199-217.
- Culver, S. M., Wolfe, L. M., & Cross, L. H. (1990). Testing a model of teaching satisfaction for Blacks and Whites. *American Educational Research Journal*, 27, 323-349.
- Dana, R. H. (1997). *Understanding cultural identity in intervention and assessment*. Thousand Oaks, CA: Sage.
- Dehue, T. (2001). Establishing the experimental society: The historical origin of social experimentation according to the randomized controlled design. *American Journal of Psychology*, 114, 283-302.

- Dornbusch, S. M., Ritter, P. L., Leiderman, H., Roberts, D. F., & Fraleigh, M. J. (1987). The relation of parenting style to adolescent school performance. *Child Development*, 58, 1244-1257.
- Duran, B. J., & Wefter, R. E. (1992). Immigrants' aspirations, high school process, and academic outcomes. *American Educational Research Journal*, 29, 163-181.
- Finn, J. D., & Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27, 557-577.
- Flynn, J. R. (1991). *Asian Americans: Achievement beyond IQ*. Mahwah, NJ: Erlbaum.
- Frisby, C. L. (2001). Academic achievement. In L. Suzuki, J. G. Ponterotto, & P. J. Meller (Eds.), *Handbook of multicultural assessment* (2nd ed., pp. 541-568). San Francisco: Jossey-Bass.
- Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Arnold.
- Gibson, M. A. (1988). *Accommodation without assimilation: Sikh immigrants in an American high school*. Ithaca, NY: Cornell University Press.
- Gopnik, A., & Meltzoff, A. N. (1992). Categorization and naming: Basic-level sorting in 18-month-olds and its relation to language. *Child Development*, 63, 1091-1103.
- Griffen-Pierson, S. (1990). The competitiveness questionnaire: A measure of two components of competitiveness. *Measurement and Evaluation in Counseling and Development*, 23, 108-115.
- Gutierrez, J., & Sameroff, A. (1990). Determinants of complexity in Mexican-American and Anglo-American mothers' conceptions of child development. *Child Development*, 61, 384-394.
- Gutierrez, J., Sameroff, A., & Karrer, B. M. (1988). Acculturation and SES effects on Mexican-American parents' concepts of development. *Child Development*, 59, 250-255.
- Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53, 87-97.
- Harter, S. (1999). *The construction of the self: A developmental perspective*. New York: Guilford Press.
- Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, 47, 1083-1101.
- Hui, C. H., & Triandis, H. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology*, 20, 296-309.
- Johnson, B. (2001). Toward a new classification of nonexperimental quantitative research. *Educational Researcher*, 30, 3-13.
- Kamin, L. J. (1974). *The science and politics of I.Q.* Mahwah, NJ: Erlbaum.
- Keefe, S. E., & Padilla, A. M. (1987). *Chicano ethnicity*. Albuquerque, NM: University of New Mexico Press.
- Kerlinger, F. N. (1979). *Behavioral research*. Austin, TX: Holt, Rinehart and Winston.
- Kim, U. (1988). *Acculturation of Korean immigrants to Canada: Psychological, demographic and behavioural profiles of emigrating Koreans, non-emigrating Koreans and Korean-Canadians*. Unpublished doctoral dissertation, Queen's University, Kingston, Ontario, Canada.
- Knight, G. P., Tein, J. Y., Shell, R., & Roosa, M. (1992). The cross-ethnic equivalence of parenting and family interaction measures among Hispanic and Anglo American families. *Child Development*, 63, 1392-1403.
- Kraft, C. L. (1991). What makes a successful Black student on a predominantly White campus? *American Educational Research Journal*, 28, 423-444.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.
- Lamborn, S. D., Mounts, N. S., Steinberg, L., & Dornbusch, S. M. (1991). Patterns of competence and adjustment among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child Development*, 62, 1049-1065.
- Landrine, H., & Klonoff, E. A. (1996). *African American acculturation*. Thousand Oaks, CA: Sage.
- Laosa, L. (1982). School, occupation, culture and the family: The impact of parental schooling on the parent-child relationship. *Journal of Educational Psychology*, 74, 791-827.
- Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity. *Educational Researcher*, 31, 3-12.
- de la Luz Reyes, M., & Halcón, J. J. (1988). Racism in academia: The old wolf revisited. *Harvard Educational Review*, 58, 299-314.
- Marchant, G. J. (1991). A profile of motivation, self-perceptions and achievement in black urban elementary children. *Urban Review*, 23, 83-99.
- Marin, G. (1993). Influence of acculturation on familialism and self-identification among Hispanics. In M. E. Bernal & G. P. Knight (Eds.), *Ethnic identity: Formation and transmission among Hispanics and other minorities* (pp. 181-196). Albany, NY: State University of New York Press.
- Marín, G., Gamba, R. J., & 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.
- Marín, G., & Marín, B. V. (1991). *Research with Hispanic populations*. Thousand Oaks, CA: Sage.
- Martínez, R., & Dukes, R. L. (1987). Race, gender, and self-esteem among youth. *Hispanic Journal of Behavioral Sciences*, 9, 427-443.
- Matsumoto, D. (1994). *Cultural influences on research methods and statistics*. Pacific Grove, CA: Brooks/Cole.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Moyerman, D. R., & Forman, B. D. (1992). Acculturation and adjustment: A meta-analytic study. *Hispanic Journal of Behavioral Sciences*, 14, 163-200.
- Nederhof, A. J. (1985). Methods of coping with social desirability: A review. *Journal of European Social Psychology*, 15, 263-280.
- Oetting, E. R., & Beauvais, F. (1990/1991). Orthogonal cultural identification theory: The cultural identification of minority adolescents. *International Journal of the Addictions*, 25, 655-685.
- Olmedo, E. (1979). Acculturation: A psychometric perspective. *American Psychologist*, 34, 1061-1070.
- Padilla, A. M. (Ed.). (1980). *Acculturation: Theory, models, and some new findings*. Boulder, CO: Westview Press.
- Padilla, A. M. (1988). Early psychological assessment of Mexican-American children. *Journal of the History of the Behavioral Sciences*, 24, 111-117.
- Padilla, A. M., & Gonzalez, R. (2001). Academic performance of immigrant and U.S. born Mexican heritage students: Effects of schooling in Mexico and bilingual/English language instruction. *American Educational Research Journal*, 38, 727-742.
- Parham, T. A., White, J. L., & Ajamu, A. (1999). *The psychology of Blacks: An African-centered perspective*. Upper Saddle River, NJ: Prentice Hall.

- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46, 598-609.
- Pedhazur, E. J., & Pedhazur-Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach* (2nd ed.). Mahwah, NJ: Erlbaum.
- Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin*, 108, 499-514.
- Phinney, J. S., & Rotheram, M. J. (Eds.). (1987). *Children's ethnic socialization*. Thousand Oaks, CA: Sage.
- Plutchik, R. (1974). *Foundations of experimental research*. New York: HarperCollins.
- Portes, A., & Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. Berkeley, CA: University of California Press.
- Richman, A. L., Miller, P. M., & LeVine, R. A. (1993). Cultural and educational variations in maternal responsiveness. *Developmental Psychology*, 28, 614-621.
- Rodríguez, C. E. (1989). *Puerto Ricans born in the U.S.A.* Boston: Unwin Hyman.
- Rogler, L. H. (1989). The meaning of culturally sensitive research in mental health. *American Journal of Psychiatry*, 146, 296-303.
- Rogler, L. H. (1999). Methodological sources of cultural insensitivity in mental health research. *American Psychologist*, 54, 424-433.
- Rogler, L. H., Cortes, D. E., & Malgady, R. G. (1991). Acculturation and mental health status among Hispanics: Convergence and new directions for research. *American Psychologist*, 46, 585-597.
- Rogler, L. H., Mroczek, D. K., Fellows, M., & Loftus, S. T. (2001). The neglect of response bias in mental health research. *Journal of Nervous and Mental Disease*, 189, 182-187.
- Rosenberg, M. (1979). *Conceiving the self*. New York: Basic Books.
- Ross, C. E., & Mirowsky, J. (1984). Socially desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior*, 25, 189-197.
- Rotenberg, K. J., & Cranwell, F. R. (1989). Self-concept in American Indian and White children. *Journal of Cross-Cultural Psychology*, 20, 39-53.
- Sears, D. O. (1986). College sophomores in the laboratory: Influence of a narrow data based on social psychology's view of human nature. *Journal of Personality & Social Psychology*, 51, 515-530.
- Shen, B., & Takeuchi, D. T. (2001). A structured model of acculturation and mental health status among Chinese Americans. *American Journal of Community Psychology*, 29, 387-418.
- Stevenson, H. W., Chen, C., & Uttal, D. H. (1990). Beliefs and achievement: A study of Black, White, and Hispanic children. *Child Development*, 61, 508-523.
- Stigler, J. W., Smith, S., & Mao, L. (1985). The self-perception of competence by Chinese children. *Child Development*, 56, 1259-1270.
- Strom, R., Johnson, A., Strom, S., & Strom, P. (1992). Educating gifted Hispanic children and their parents. *Hispanic Journal of Behavioral Sciences*, 14, 383-393.
- Suarez, Z. E. (1993). Cuban Americans: From golden exiles to social undesirables. In H. P. McAdoo (Ed.), *Family ethnicity: Strength in diversity* (pp. 164-176). Thousand Oaks, CA: Sage.
- Suarez-Orozco, M. M. (1989). *Central American refugees and U.S. high schools: A psychosocial study of motivation and achievement*. Stanford, CA: Stanford University Press.
- Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist*, 54, 1070-1077.
- Sue, S., & Okazaki, S. (1990). Asian American educational achievements: A phenomenon in search of an explanation. *American Psychologist*, 45, 913-920.
- Suinn, R. M., Richard-Figueroa, K., Lew, S., & Vigil, P. (1987). The Suinn-Lew Asian self-identity acculturation scale: An initial report. *Educational and Psychological Measurement*, 47, 401-407.
- Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students: Foundations, performance factors, and assessment issues*. Thousand Oaks, CA: Sage.
- Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, 62, 1066-1078.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Wylie, R. (1979). *The self-concept: Vol. 2. Theory and research on selected topics*. Lincoln, NB: University of Nebraska Press.
- Yee, A. H. (1992). Asians as stereotypes and students: Misperceptions that persist. *Educational Psychology Review*, 4, 95-132.
- Yoder, J. D., & Kahn, A. S. (1993). Working toward an inclusive psychology of women. *American Psychologist*, 48, 846-850.

Yet multicultural education is an international discipline and field of study. It takes unique forms in each nation, yet it shares a number of overarching issues, concepts, and paradigms cross-nationally. The chapters in Part XII describe the historical development and current status of multicultural education research and developments in Australia, the United Kingdom, and South Africa. His claim applies equally well to the second edition of the Handbook of Research on Multicultural Education, ably edited by James A. Banks and Cherry A. McGee Banks. Newly revised and expanded, this volume indeed contains a multitude of voices and perspectives, providing an invaluable resource for all those interested in issues of equity and the role of education in a pluralistic democracy.