

---

# Science, Ethnicity, and Bias

---

## *Where Have We Gone Wrong?*

---

Stanley Sue  
University of California, Davis

*The quality, quantity, and funding of ethnic minority research have been inadequate. One factor that has contributed to this inadequacy is the practice of scientific psychology. Although principles of psychological science involve internal and external validity, in practice psychology emphasizes internal validity in research studies. Because many psychological principles and measures have not been cross-validated with different populations, those conducting ethnic minority research often have a more difficult time demonstrating rigorous internal validity. Thus, psychology's overemphasis of internal as opposed to external validity has differentially hindered the development of ethnic minority research. To develop stronger research knowledge on ethnic minority groups, it is important that (a) all research studies address external validity issues and explicitly specify the populations to which the findings are applicable; (b) different research approaches, including the use of qualitative and ethnographic methods, be appreciated; and (c) the psychological meaning of ethnicity or race be examined in ethnic comparisons.*

One issue has bedeviled psychology for many decades—namely, the relationship between racial and ethnic bias and the practice of psychological science. In 1972, Thomas and Sillen discussed “scientific racism” and the ways in which theories and empirical research perpetuated a biased view of African Americans and other ethnic minority groups (Thomas & Sillen, 1972). Indeed, in the foreword to their book, Kenneth B. Clark, past president of the American Psychological Association, noted that

Probably the most disturbing insight obtained from the relentless clarity with which this book documents the case of racism . . . is the ironic fact that the students, research workers, and professionals in the behavioral sciences—like members of the clergy and educators—are no more immune by virtue of their values and training to the disease and superstition of American racism than is the average man. (p. xii)

Most of the controversies presented in the book are still being debated in the field. For example, issues over race and intelligence, test bias, equivalence of measures (conceptual, translation, and scalar equivalence) across different ethnic groups, deficit model theories, the design and interpretation of research on ethnic minority groups, stereotyping, focus on social structure versus the individual,

and the influence of culture continue to be prominent and contentious themes in American psychology (Herrnstein & Murray, 1994; Ivey, Ivey, & Simek-Morgan, 1997; Jones, 1997; Ponterotto & Pedersen, 1993; Sandoval, Frisby, Geisinger, Scheuneman, & Greiner, 1998; Segall, Lonner, & Berry, 1998; Steele, 1997; D. W. Sue & Sue, 1999; Triandis & Brislin, 1984).

The questions I address are these: Is science biased against ethnic minority research? Are ethnic minority researchers simply playing, as some believe, a “White man’s game”? Should ethnic minority researchers turn to alternatives rather than traditional experimental and empirical methods? Why is there a relative lack of ethnic minority research and high-quality ethnic research? How can we promote and encourage good ethnic minority research?

Let me state from the outset my position: I believe that there is a lack of psychological research on ethnic minority populations; that research on ethnic minority groups is uneven, with much of it at a relatively low level; and that funding for ethnic minority research has been woefully inadequate. The reasons for these problems are subtle and systemic. Science and scientific methods are not the culprit. Rather, the culprit is how science has been practiced—an effect caused by the selective enforcement of the principles of science. This selective enforcement of the principles of science emphasizes internal validity over external validity, which discourages the growth and development of ethnic minority research. As a consequence, steps must be taken in psychology to rectify the situation, not only in the interest of ethnic minority populations but also in the promotion of better science. (Incidentally, I realize terms such as *Whites*, *Asian Americans*, *ethnic minority group*, etc., are often offensive, stereotypic referents. My use of the terms refers to the aggregate group rather than to individuals.)

---

An earlier version of this article was presented as a keynote address at the National Multicultural Conference and Summit, Newport Beach, California, January 1999. The article is supported in part by Grant R01-MH44331 from the National Institute of Mental Health. I am indebted to Sumie Okazaki, Rick Robins, and David Takeuchi for their comments on this article.

Correspondence concerning this article should be addressed to Stanley Sue, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616-8686. Electronic mail may be sent to [ssue@ucdavis.edu](mailto:ssue@ucdavis.edu).



**Stanley Sue**

## **Ethnic Research**

Having served on editorial boards and grant review panels, I have often encountered colleagues who criticize much of ethnic minority research as being descriptive in nature, simple in design, and lacking in grounding in sophisticated and mature psychological theory. These are, of course, sufficient justification for not accepting a paper for publication or for rejecting proposals for funding. Even in my own research, I have often resorted to basic demographic and descriptive studies. For example, I examined the kinds of problems encountered by Asian Americans, who are often perceived as a model minority group, and the correlates of their utilization of mental health services (S. Sue, Sue, Sue, & Takeuchi, 1995). I felt it was important to point to ethnic differences in behaviors, and these studies added new knowledge to the field. However, the nature of the knowledge was basic and descriptive about Asian Americans. From the data, it was difficult to draw inferences concerning the reasons for the utilization rates or the implications for other populations. Therefore, the work had value in understanding one particular ethnic minority group but unclear value in generalizing beyond that group. The discrepancy in values also meant that research of this kind could engender different reactions from reviewers. One reviewer might view the research positively because of an understanding of the nature of the ethnic field and an ability to place the value of the contribution in context. Another reviewer might negatively evaluate the same work because of a lack of understanding of the field, a desire to have the work demonstrate greater generality than the targeted group, or uncertainty as to the theoretical rationale for the research. In cases where reviewers disagree about the merit of a research paper submitted for publication, the typical practice is to not accept the paper for publication.

There is evidence that the journal literature contains relatively few articles on ethnic minority groups. For example, Graham (1992) examined the number of empirical research articles on African Americans in several psychology journals. She identified articles in which African Americans were the population of interest or in which data were analyzed by race and included African Americans. These articles constituted only 3.6% of the total published between 1970 and 1989. Surprisingly, there was actually a decline in the number of such articles over time. After conducting a content analysis of three behavioral psychology journals for empirical or theoretical and treatment-oriented or assessment-oriented articles, Iwamasa and Smith (1996) found only 1.3% of the articles focused on U.S. ethnic minority groups. Padilla and Lindholm (1995) also noted a paucity of publications existed on topics devoted to race, ethnicity, and culture in major psychology journals.

One parsimonious explanation for the state of ethnic minority research is that there are few researchers interested in the topic and that the ethnic field is relatively underdeveloped. In this view, it will take time before culturally appropriate research measures, tools, and methodologies can be established. I do not believe that the notion of underdevelopment can fully explain the disparity in research sophistication and publication rates. Rather, the current practices in psychological science have contributed to the problem. To understand how this is possible, we need to examine the philosophy of science, epistemology, and research principles.

## **Science**

There are many ways of knowing. Widely recognized philosophers from Descartes to Kant to Wittgenstein have tried to identify how one knows, what one knows, and what it means to know. Science is a way to know things. There are other ways of knowing, such as having spiritual awareness or personal intuition (a form of knowledge or of cognition independent of experience or reason). These ways of knowing cannot be distinguished from each other simply on the basis of being correct or incorrect. Knowledge generated from science as well as spiritual and intuitive means can be wrong. For example, in science, we can draw invalid conclusions not only because of methodological and conceptual problems but also because of scientific convention. When alpha levels are specified in testing hypotheses, Type I and Type II errors will inevitably occur in the long run.

Science is used to describe, explain, predict, and modify phenomena. There are several basic guiding principles in the philosophy of science, including convincingness (credibility) and skepticism. The principle of convincingness refers to the direct relationship between the use of rigorous scientific methodology and the likelihood of convincing or persuading a fair-minded, critically thinking person of the findings or conclusions. Through the elegance and soundness of scientific methodology and theoretical coherence, one gains confidence that the findings are valid (Abelson, 1995); the more rigorous the study, the better.

The continuing efforts to refine science and to design rigorous studies have led to the use of operational definitions, replication, matching, manipulation checks, control groups, random assignments, experimental procedures, blind studies, sampling techniques, theory building, methods of controlling confounding variables, follow-ups, statistical power, clinical significance, sophisticated statistical procedures, and the like.

Another important principle is skepticism. Under this principle, one must question, doubt, or suspend judgment until sufficient information is available. Skeptics demand that evidence and proof be offered before conclusions can be drawn (Marken, 1981). Scientific skepticism is considered good. One must thoughtfully gather evidence and be persuaded by the evidence rather than by prejudice, bias, or uncritical thinking. The basis for the principle comes from philosophers and scientists who embrace logical positivism and logical empiricism.

In psychological science, one of the most critical tasks to master is the ability to draw causal inferences from research. We want to understand why something occurs, that is, the causal factors. To do this, the research must have internal and external validity. Internal validity concerns the extent to which conclusions can be drawn about the causal effects of one variable on another (i.e., the effect of the independent variable on the dependent variable). The interest in making causal inferences goes beyond quantitative research. Even in qualitative research, a variety of strategies are used to strengthen internal validity so causal inferences can be drawn, such as the use of prolonged substantial engagement, persistent observations, peer debriefing, negative case analysis, progressive subjectivity, member checks, and triangulation (Mertens, 1998).

External validity is the extent to which one can generalize the results of the research to the populations and settings of interest. One tries to study the population of interest. If sampling is necessary, a variety of strategies can be used to generate a sample that is representative of the population. There are some limited circumstances in which generality of findings is not always relevant or necessary (Kazdin, 1999). However, more typically, both internal and external validity are important and should be equal partners in the scientific endeavor. The lack of internal validity does not allow causal inferences to be made without some degree of convincingness or credibility. The lack of external validity may render findings meaningless with the actual population of interest. A natural tension always exists between internal and external validity. For experimenters to achieve perfect internal validity, a laboratory is often the perfect setting. For experimenters to achieve perfect external validity, research is conducted in the actual environment with the accompanying noise and extraneous variables that normally exist.

### **Selective Enforcement of Scientific Principles**

It is my belief that internal and external validity are not equal partners, and that if we have erred, it is in the direction of attention to internal rather than external valid-

ity in psychological science. Far more papers submitted for publication are probably rejected because of design flaws, inability to control for confounding or extraneous variables, and so on than for possible external validity problems. Because one must first obtain an unambiguous finding before one can generalize about it, internal validity often takes precedence over external validity (Alloy, Abramson, Raniere, & Dyller, 1999). Indeed, one could argue that the demand for rigor in internal validity far exceeds that for external validity. Consider the fact that college student samples are used extensively in psychological research without much attention to the representativeness or generality of the sample. Kazdin (1999) noted that students are typically enrolled in undergraduate psychology courses and receive monetary rewards or course credit for participation. Is this sample representative of human beings in general or college students in particular? Sears (1986) found that the use of college sophomores as participants has resulted in a biased view of human beings that reflects the idiosyncrasies of college students. Even the widely used practice of selecting (self-reported) distressed college students to study depression has been criticized as being inappropriate (Coyne, 1994). When college students are used, ethnicity or race is typically ignored as a variable of interest, and the underrepresentation of certain minority groups and persons from different cultures as research participants is of concern (Kazdin, 1999).

Another fact should be considered. Americans are the largest producers of psychological research. The overwhelming subject of the research is Americans. The United States constitutes less than 5% of the world's population. Therefore, on the basis of a sample of less than 5% of the world, theories and principles are developed that are mistakenly assumed to apply to human beings in general; that is, they are assumed to be universal. Seldom are researchers only interested in whether an independent variable affects the dependent variable for the particular participants used in the investigation. They usually want to make a general statement about some larger population (Robinson, 1976). They then make the mistake of discussing human beings without much reference to the particular sample on which the research was based.

Interestingly, Korchin (1980) noted that researchers tend to question the generality of findings only when the research involves ethnic minority populations. Korchin mentioned that once he and his colleague had conducted research on why some African American youths had made extraordinary achievements. A paper from the research was submitted for publication and rejected. One reviewer had indicated that the research was grievously flawed because it lacked a White control group. Why was a White control group necessary if the interest was in African Americans? More critically, Korchin asked why we do not require studies of Whites to have an African American control group. In other words, we ask that ethnic minority research show its pertinence to other groups or more general phenomenon, but we fail to make the same requests when the research involves White populations. The answer to the

question of when a control group is needed depends on the research question being asked.

To further illustrate the tendency to assume generality, let me describe my experiences as a member of the APA Division 12 Task Force on Psychological Interventions. The charge of the task force was to determine which psychotherapeutic interventions had met rigorous criteria (including replication) for being designated an empirically validated treatment. During the course of the investigation, the task force (Chambless et al., 1996) found that no rigorous studies had ever been conducted on the effectiveness of treatment for members of ethnic minority groups. How, then, can treatments be designated as validated, where validity is attributed to the treatment itself, when cross-cultural validity has never been established? The dilemma was recognized by researchers, and the phrase "empirically supported treatment" rather than "empirically validated treatment" was subsequently adopted (Chambless & Hollon, 1998). Therefore, this situation shows that the practice in research has often been to assume generality of findings when it is not warranted, an external validity problem. Abelson (1995) indicated that generality of effects is often limited by cultural factors and that universals are rare among human beings.

The tendency to ignore external validity problems (which can also impact internal validity) is also illustrated by contrasts made by Padilla and Lindholm (1995) in their analysis of descriptions of participant populations in research studies. For example, these extracts were in two different research reports Padilla and Lindholm cited:

Subjects were 32 children from the Berkeley area. Subjects included 15 boys and girls. Children came from a variety of socioeconomic backgrounds, though most were Caucasian and middle class. (p. 100)

In all, 423 sixth- and seventh-grade students . . . participated in the study. The school was in a predominantly working-class . . . 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. (p. 100)

Padilla and Lindholm noted that in these two descriptions there was basic demographic information, but it was too general to be meaningful and not cross-tabulated. For example, in the second study, how many sixth graders were girls? How many of the boys were members of various ethnic groups? Without such information, it would be difficult to replicate the study. In contrast, the following description was made of an experimental study involving primates:

The subjects were 28 pigtailed macaque infants. Of these, 17 were classified as normal or at low risk for developmental deficits (i.e., normal birth weight, no history of clinical problems), and 11 were classified as high risk. Of the high-risk animals, 6 had experienced significant trauma at birth or during the early neonatal period (e.g., breech delivery with respiratory complications). Three high-risk subjects were part of an ongoing study examining the effects of parental reproductive history and maternal stress on growth and

development. . . . The infants were separated from their mothers at birth. . . . All animals were socialized in play groups daily. (as quoted in Padilla & Lindholm, 1995, p. 101)

The detailed information reflects the importance that the researchers placed on the sample. As mentioned earlier, in psychological research with human beings, we seem to deemphasize background characteristics and to draw broad conclusions about human beings in general rather than about the particular human beings in the study. In other words, our *modus operandi* is to assume that the work is universally applicable; the burden of proof is placed on researchers concerned about race, ethnicity, and bias to show that there are ethnic differences. Whatever happened to the scientific notion of skepticism, where little is taken for granted, where conclusions are drawn from evidence and not from assumptions? We have not followed good scientific principles in assuming that findings from research on one population can be generalized to other populations. At a recent meeting of psychology journal editors, the need for better demographic descriptions of study participants, including more precise definitions of race and ethnicity, was emphasized (Azar, 1999).

The phenomenon of selective enforcement of scientific criteria is apparent. We criticize research for problems in internal validity. Yet, we pay relatively little attention to external validity. Whether in research papers submitted for publication or research grant proposals submitted for funding, ethnic minority research is primarily scrutinized for internal validity problems. This is appropriate. Yet, much of research, whether or not it is focused on ethnics, is not criticized for external validity problems. In experimental studies, the discrepancies between internal and external validity are brought into bold relief. Rigorous and sometimes elegant experimental designs allow us to make causal inferences. However, because of their rigor and need to control for extraneous variables, they frequently involve small numbers of participants, foregoing issues of sampling and representativeness. The ultimate example of this is the single participant experimental design. In dealing with a single participant experiment, such as an ABAB time-series design, extraneous variables like maturation and reactivity can be controlled, and, in the process, alternative explanations can be eliminated. The generality of findings from such a design is unclear.

A similar situation occurs in the mental health field, where a distinction is made between efficacy and effectiveness research. On the one hand, in the past, efficacy studies were considered the gold-standard methodology in evaluating the outcome of mental health treatments. These studies were important in demonstrating the causal effects of a particular intervention on mental health symptoms. However, experimenters tended to use highly restrictive exclusion criteria that enhanced internal validity in their designs. For example, in an efficacy study of a psychotherapeutic treatment for schizophrenia, exclusionary criteria might include comorbidity, current use of psychotropic medication, and non-White race. Such a practice would limit the applicability or generality of the findings. On the other

hand, effectiveness research is gaining appreciation because it examines treatment outcomes in more typical real-life situations, where patients may have multiple disorders, be taking medication, and be members of various racial groups.

### **Consequences of the Selective Enforcement of Scientific Principles**

Focus on internal validity has had the effect of discouraging ethnic minority research. Because much psychological research is not based on ethnic minority populations, it is unclear whether a particular theory or principle is applicable to all racial groups, whether an intervention has the same phenomenological meaning for different cultural groups, or whether measures or questionnaires are valid for these populations. Ethnic minority research has a weaker substantive base on which to demonstrate internal validity. Consequently, much effort has to be expended in discovering basic relationships between variables, developing measures, and so on. For example, ethnic researchers cannot simply assume that Minnesota Multiphasic Personality Inventory—2 is a valid instrument for, say, Cambodian Americans. Much effort must be expended in translating, backtranslating, and testing the instrument. Ironically, in some cases, by the time an instrument originally designed for White Americans has been found to be valid for an ethnic minority population, a new, revised version is available for White Americans, and another round of validation must be initiated for the ethnic minority population.

The greater the demands for internal validity, the more the mainstream research will dominate. To be sure, the demand for rigor in internal validity will improve ethnic minority research, but the gap will widen between ethnic minority and mainstream research. A vicious cycle is created in which demands for rigorous internal validity result in relatively few minority research projects and less funding; the reduction in research and funding makes it difficult to develop appropriate measures, designs, and research tools; training opportunities in ethnic minority research are difficult to find; it is then a problem to meet demands for rigor. It should be noted that I am not trying to characterize all ethnic minority research as lacking rigor and internal validity. There are many examples of outstanding, rigorously designed studies conducted on ethnic minority populations. My main point is there are systemic forces that discourage the development of ethnic minority research. Unfortunately, these forces are not well recognized, and much ethnic minority research has been negatively viewed.

Another factor that discourages ethnic minority research is that such research is difficult to conduct. Researchers may avoid ethnic minority research for a variety of reasons, including difficulties in recruiting ethnic participants, not having validated measures to use, the unknown validity of mainstream theories, and the often controversial nature of ethnic minority research (e.g., the heated debates over race and intelligence). As a case in point, several of us

at our National Institute of Mental Health-funded research center were interested in conducting the first large-scale study of the rate, distribution, and correlates of mental disorders among Chinese Americans. In previous national surveys of the mental health of Americans (Kessler et al., 1994; Robins & Regier, 1991) Chinese Americans and Asian Americans were not analyzed, presumably because of their small sample sizes in the studies. Because of the problems in finding a representative sample of a relatively small minority group, devising valid measures and translating instruments into Chinese, training bilingual and bicultural interviewers, and so on, it cost \$1.5 million to conduct a two-wave study (see Takeuchi et al., 1998). Nearly 20,000 households were approached to find 1,700 Chinese respondents. Thus, ethnic minority research can be very costly and difficult to initiate. Given the difficulties in conducting research, the field's emphasis on internal validity, and the paucity of ethnic minority studies, it is not surprising that ethnic minority research continues to lag behind that of the mainstream.

My position is not that psychological principles or theories cannot be generalized from one population to another. Many principles can be applied to different populations. Problems occur when the assumption of generality is made. Generality is a phenomenon that should be empirically tested. It is in the best interest of science, and it is certainly consistent with the tradition of scientific skepticism, to avoid drawing premature and untested assumptions. Moreover, when theories and models applied to different populations are examined, important ethnic and cultural differences are often found.

### **Ethnic and Cultural Differences: Two Examples**

Research on ethnic differences has yielded important insights about the applicability of theories and models. For example, my colleagues and I have examined ethnic differences in sexual aggression (Hall, Sue, Narang, & Lilly, in press). Theoretical models of men's sexual aggression have emphasized intrapersonal factors (Hall & Barongan, 1997): the physiological, cognitive, affective, and developmental variables that may motivate them to become sexually aggressive. Such perpetrator characteristics are more strongly associated with sexual aggression than are victim characteristics. After removing the effects of social desirability, intrapersonal variables (rape myth acceptance, hostility toward women, and the belief that rape is morally wrong) accounted for a significant increment in the amount of variance observed in sexual aggression. Situational variables (number of consenting sex partners, drinking before or during sex, being drunk before or during sex) did not account for a significant amount of variance in sexual aggression.

Whereas the motivation for behavior in many Western cultures tends to be individually based or independent in nature, motivation for behavior in collectivist cultures tends to depend on social context or be interdependent (Markus & Kitayama, 1991). Unlike the emphasis on individualism in mainstream American culture, American cultures having non-Western origins often emphasize col-

lectivist values. Collectivist cultures emphasize attending to others, fitting within the community, role relationships, and interpersonal harmony.

We hypothesized that specific aspects of collectivist cultures may serve as protective factors against sexually aggressive behavior. The importance of fitting in with the group may cause group norms to be more influential in collectivist than in individualist cultures. We conducted two studies in which we assessed self-reported sexual aggression, intrapersonal factors (rape myth acceptance and hostility toward women), and interpersonal factors (concerns over loss of face and effects of sexual aggression on relationships with parents) among Asian American and White male university students. The results of the two studies supported the hypothesis that interpersonal variables would be more relevant in a model of sexual aggression for Asian American men than in a model for White men. Although intrapersonal variables were important for both groups, interpersonal variables were a predictor of sexual aggression only for Asian Americans. The preliminary findings on sexual aggression may well imply fundamental differences in explanatory models for individualistic and collectivist cultures.

In another study, we were interested in possible ethnic differences in predictors of academic achievement. Past research has demonstrated the importance of high school grades and Scholastic Aptitude Test (SAT) scores (particularly the verbal portion of the SAT) in predicting subsequent grades in college. Because little research had been conducted on Asian Americans, we wanted to examine the predictors for this population (Sue & Abe, 1988).

From a total freshman student population of 22,105 who enrolled in any of the eight University of California campuses, the records of the 4,113 domestic Asian (non-foreign) students were examined and compared with those of 1,000 randomly selected White students. Asian American students were also divided into two groups by presumed English proficiency: those for whom English was probably the best language and those for whom English was probably not the best language. The criterion variable was university freshman grade point average (GPA), which was the average of all grades received by a student during the academic year. Six predictor variables were used for the GPA: high school GPA, SAT-Verbal (SAT-V) score, SAT-Mathematical (SAT-M) score, English composition score from the College Board Achievement Test (CBAT) series, and Level I or Level II mathematics test (MI or MII) score from the CBAT series. Because the overall results with the SAT scores were similar to those using the English composition and the MI and MII CBAT scores, only the results for high school GPA and SAT scores are presented.

Very different regression formulas best predicted university GPA for Whites and Asian Americans. For Whites, high school GPA and SAT-V were the best predictors of university GPA, with SAT-M adding virtually nothing to the prediction equation. For Asian Americans, high school GPA and SAT-M were the best predictors, with SAT-V adding very little insight. The results were consistent regardless of declared academic major, gender, or accultur-

ation level of the students. Interestingly, important differences within the Asian American sample were found, depending on precise ethnicity. The results are important because in test bias research, one indicator of bias is whether regression slopes for predicting behaviors differ for different groups. We have evidence that they do—culture and ethnicity do make a difference.

As noted by Kazdin (1999) and Lin, Poland, and Nakasaki (1993), even biological responses to psychotropic medication (e.g., for depression and anxiety) can vary greatly as a function of racial and ethnic differences. "Such findings underscore the need to sample broadly, to evaluate the moderating role of sample differences, and to pursue mechanisms through which moderating factors may operate" (Kazdin, 1999, p. 20). These findings provide convergent evidence for the importance of testing the validity of models and theories with different populations. The findings inevitably raise the question of whether all theories must be tested with every population, surely an impossible task. My response is "yes." We cannot assume *a priori* that explanatory models apply to different populations in the absence of evidence or logical reasoning. My position is consistent with the principle of scientific skepticism.

## Some Solutions

How do we go about increasing the quality and quantity of research on ethnic minority populations? Interestingly, the National Institute of Mental Health has attempted to encourage researchers to include female and ethnic minority research participants (Hohmann & Parron, 1996) by issuing several research guidelines:

1. Women and members of minorities and their subpopulations must be included in all human participant research.
2. Cost is not an acceptable reason for excluding these groups.
3. Programs and support for outreach efforts should be initiated to recruit these groups into clinical studies.

Because of the guidelines, many researchers complained that they were often being compelled to modify their projects. They had to justify why they did not include women and minorities in their studies. Some attributed the policy to "political correctness." Hohmann and Parron argued that the policy reflected scientific, not political, concerns. The scientific principle was based on the heterogeneity of the population of the United States, which necessitated research that would reflect the population's diversity.

Both the research complainants and Hohmann and Parron (1996) were correct to some degree. On the one hand, of course there were political considerations. The policy would not have been adopted if there were not political considerations and pressures. On the other hand, critics of the policy failed to see the scientific merit of inclusion. The fact that the nation is heterogeneous makes the inclusion policy important. However, heterogeneity *per se* is not a scientific matter. It is when we generalize from



one sample to other populations that a scientific principle is pertinent. This is the external validity issue. If we want to draw conclusions about human beings, we must study human beings in all their diversity and not particular samples.

It should be noted that engaging in ethnic minority research does not magically induce external validity. A focus on different populations can be used to discover how universal research findings or theories are. Ethnicity should not be treated as a nuisance variable. Understanding ethnic differences is not only helpful to ethnic groups, it is good for science. The United States is one of the most diverse societies in the world. Why not take advantage of that fact by promoting external validity and by testing the generality of theories?

Interestingly, psychologists in the testing and assessment field have taken several major steps to emphasize the importance of the external validity of instruments. The principles they propose researchers follow (see Sandoval et al., 1998) include an explicit statement of the nature of the sample on which tests have been standardized and the population(s) for which a test is appropriate. They also note that it is incorrect to consider a test valid for all populations because no test is valid for all purposes or in all situations.

I believe that progress in the development of ethnic minority research in particular and research in general will be well served by engaging in several tasks. First, ethnic minority research will be improved if demands are made for greater external validity while strong internal validity standards are maintained in all research studies. There should also be a true desire to include many different populations in research in order to find out if phenomena are universal or particularistic. Those who construct theories and principles in psychology ought to explicitly detail the population on which the theories and principles are built. In cases where those theories and principles are built on limited populations, they should be viewed as local theories until they are cross-validated. The intent of this recommendation is not to imply that theories and principles lack generality: They may or may not. The point is that they require evidence, in the best tradition of science and the principle of skepticism, and cross-validation. The burden of proof is to show that results are universally applicable. By explicitly delineating the limitations in generality, researchers may be motivated to study other populations to establish the applicability of principles and theories. Similarly, in ethnic minority research, it is important to not only gain knowledge of the particular targeted ethnic group but also contribute to the understanding of more general processes, principles, or limitations in these principles.

Second, a wide range of research methodologies should be used, particularly in ethnic minority research. Traditional psychological scientists tend to prefer experimental methods in the study of cause and effect. As mentioned previously, experimenters conducting ethnic minority research often cannot assume that existing theories or propositions developed on other populations are strictly applicable to the population they are studying. Therefore, a

variety of research methods, including qualitative and ethnographic strategies, should be used to create a more substantial knowledge base. Sears (1986) has suggested that very different kinds of people in very different behavioral settings using different research methodologies should be recruited as participants to avoid the study of "college sophomores in the laboratory."

Third, ethnicity is a distal and largely demographic concept in explaining observed differences. Ethnic and cross-cultural researchers should try to go beyond comparisons of different populations. If ethnic differences are found, researchers should explain what aspects of ethnicity are responsible for the differences. For example, as noted before, we found differences between Asians and White Americans in predictors of sexual aggression. Ethnicity (Asian or White) is a distal concept in explaining the differences. A more proximal explanatory concept is that Asians and Whites differ in interpersonal-intrapersonal orientation. Therefore, the meaning of ethnicity is frequently the most important variable to study.

In conclusion, science is not biased against ethnic minority groups. Rather, the problem is that criteria used to judge the suitability of research for publication and for funding are selectively enforced so that internal validity is elevated over external validity. In addition, because ethnic minority research is not as well developed as mainstream research and is more difficult to conduct, internal validity criteria suppress research progress. This in turn keeps ethnic minority research from becoming better developed. Ethnic minority researchers should not turn away from science; rather, demands should be made that all research show relevance for different groups. The issue of internal and external validity may also explain why psychology has been so reluctant to recognize culture. Segall, Lonner, and Berry (1998) believe that culture is too encompassing, so it is ignored or taken for granted. Another explanation is that the very tools we use as psychological scientists—research methodologies and emphases—steer us away from culture. In practice, internal validity takes precedence over the generality or specificity of findings.

## REFERENCES

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Erlbaum.
- Alloy, L. B., Abramson, L. Y., Raniere, D., & Dyller, I. M. (1999). Research methods in adult psychopathology. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (2nd ed., pp. 466–498). New York: Wiley.
- Azar, B. (1999, February). Journal editors find consortium helps researchers shape goals. *APA Monitor*, 20.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett-Johnson, S., Pope, K. S., Crits-Christoph, P., Baker, M., Johnson, B., Woody, S. R., Sue, S., Beutler, L., Williams, D. A., & McCurry, S. (1996). An update on empirically validated therapies. *Clinical Psychologist*, 49, 5–18.
- Coyne, J. C. (1994). Self-reported distress: Analog or ersatz depression? *Psychological Bulletin*, 116, 29–45.
- Graham, S. (1992). "Most of the subjects were White and middle class": Trends in published research on African Americans in selected APA journals, 1970–1989. *American Psychologist*, 47, 629–639.
- Hall, G. N., & Barongan, C. (1997). Prevention of sexual aggression:

- Sociocultural risk and protective factors. *American Psychologist*, 52, 5-14.
- Hall, G. N., Sue, S., Narang, D. S., & Lilly, R. S. (in press). Culture-specific models of men's sexual aggression: Intra- and interpersonal determinants. *Cultural Diversity and Ethnic Minority Psychology*.
- Herrnstein, R. E., & Murray, C. A. (1994). *The bell curve*. New York: Free Press.
- Hohmann, A. A., & Parron, D. L. (1996). How the new NIH guidelines on inclusion of women and minorities apply: Efficacy trials, effectiveness trials, and validity. *Journal of Consulting and Clinical Psychology*, 64, 851-855.
- Ivey, A. E., Ivey, M. B., & Simek-Morgan, L. (1997). *Counseling and psychotherapy: A multicultural perspective*. Boston: Allyn & Bacon.
- Iwamasa, G. Y., & Smith, S. K. (1996). Ethnic diversity in behavioral psychology: A review of the literature. *Behavior Modification*, 20, 45-59.
- Jones, J. M. (1997). *Prejudice and racism*. San Francisco: McGraw-Hill.
- Kazdin, A. E. (1999). Overview of research design issues in clinical psychology. In P. C. Kendall, J. N. Butcher, & G. N. Holmbeck (Eds.), *Handbook of research methods in clinical psychology* (pp. 3-30). New York: Wiley.
- Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. *Archives of General Psychiatry*, 51, 8-19.
- Korichin, S. J. (1980). Clinical psychology and minority problems. *American Psychologist*, 35, 262-269.
- Lin, K. M., Poland, R., & Nakasaki, G. (Eds.). (1993). *Psychopharmacology and the psychobiology of ethnicity*. Washington, DC: American Psychiatric Press.
- Marken, R. (1981). *Methods in experimental psychology*. Monterey, CA: Brooks/Cole.
- Markus, H. R., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98, 224-253.
- Mertens, D. M. (1998). *Research methods in education and psychology*. Thousand Oaks, CA: Sage.
- Padilla, A. M., & Lindholm, K. J. (1995). Quantitative educational research with ethnic minorities. In J. A. Banks & C. A. McGee-Banks (Eds.), *Handbook of research on multicultural education* (pp. 97-113). New York: Macmillan.
- Ponterotto, J. G., & Pedersen, P. B. (1993). *Preventing prejudice: A guide for counselors and educators*. Newbury Park, CA: Sage.
- Robins, L. N., & Regier, D. A. (1991). *Psychiatric disorders in America: The epidemiologic catchment area study*. New York: Free Press.
- Robinson, P. W. (1976). *Fundamentals of experimental psychology: A comparative approach*. Englewood Cliffs, NJ: Prentice Hall.
- Sandoval, J., Frisby, C. L., Geisinger, K. F., Scheuneman, J. D., & Grenier, J. R. (1998). *Test interpretation and diversity: Achieving equity in assessment*. Washington, DC: American Psychological Association.
- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, 51, 515-530.
- Segall, M. H., Lonner, W. J., & Berry, J. W. (1998). Cross-cultural psychology as a scholarly discipline: On the flowering of culture in behavioral research. *American Psychologist*, 53, 1101-1110.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Sue, D. W., & Sue, D. (1999). *Counseling the culturally different*. New York: Wiley.
- Sue, S., & Abe, J. (1988). *Predictors of academic achievement among Asian American and White students*. New York: College Board.
- Sue, S., Sue, D. W., Sue, L., & Takeuchi, D. T. (1995). Psychopathology among Asian Americans. *Cultural Diversity and Mental Health*, 1, 39-54.
- Takeuchi, D. T., Chung, R. C., Lin, K. M., Shen, H., Kurasaki, K., Chun, C., & Sue, S. (1998). Lifetime and twelve-month prevalence rates of major depressive episodes and dysthymia among Chinese Americans in Los Angeles. *American Journal of Psychiatry*, 155, 1407-1414.
- Thomas, A., & Sillen, A. (1972). *Racism and psychiatry*. New York: Brunner/Mazel.
- Triandis, H. C., & Brislin, R. W. (1984). Cross-cultural psychology. *American Psychologist*, 39, 1006-1016.