Improving Intergroup Relations in Higher Education: A Critical Examination of the Influence
of Educational Interventions on Racial Bias
Author(s): Mark E. Engberg
Source: *Review of Educational Research*, Vol. 74, No. 4, (Winter, 2004), pp. 473-524
Published by: American Educational Research Association
Stable URL: http://www.jstor.org/stable/3515979
Accessed: 12/08/2008 12:16

Improving Intergroup Relations in Higher Education: A Critical Examination of the Influence of Educational Interventions on Racial Bias

Mark E. Engberg
University of Michigan

This study examines the influence of various educational interventions in higher education on students' racial bias. The author reviews studies in four principle domains: multicultural courses, diversity workshops and training, peer-based interventions, and service-based interventions. He pays particular attention to the varied approaches, measures, and research designs used to assess the effectiveness of interventions. He concludes with specific recommendations for improving the quality of intervention studies, suggests a conceptual model for explaining student change, and points out gaps in the extant knowledge base. An appendix provides an overview of racial bias measures used in the reviewed studies.

KEYWORDS: diversity, higher education, intergroup relations, intervention, race.

The history of intergroup relations on college and university campuses is deeply embedded in the changing demographic composition of the postsecondary student body. Students of color, for instance, began entering colleges and universities in unprecedented numbers after the passage of the Civil Rights and Higher Education Acts in the 1960s. Yet few colleges and universities were prepared for the inherent challenges in educating such a diverse population of college students. As a result, the history of intergroup relations is marked by periods of campus unrest and heated, if not violent, exchanges between diverse groups of students (Astin, Astin, Bayer, & Bisconti, 1997).

As the nation passes the 50th anniversary of the 1954 *Brown vs. Board of Education* decision, we are reminded that desegregation and the elimination of de jure discrimination are not panaceas for improving intergroup relations. In fact, evidence in the post-*Brown* era points to increases in bias-related incidents, ranging from verbal and physical intimidation to the use of degrading and insensitive stereotypes (Dalton, 1991a). Dalton suggests a number of factors that have contributed to the rise of racial incidents: lack of knowledge, experience, and contact with diverse peers; peer-group influence; increased competition and stress; the influence of off-campus groups and the media; alcohol use; changing values; fear of diversity; and the perception of unfair treatment. Many of these factors are exacerbated

by the dearth of opportunities for students of different racial groups to have meaningful discussions about interracial issues.

Today, many campuses continue to witness increased levels of racial strife (Pettigrew, 1998a) and racial mircoaggressions (Solorzano, Ceja, & Yosso, 2000). Consequently, students of color are enmeshed in a psychological climate marked by increasing alienation (Cabrera & Nora, 1994) and detachment (Hurtado, Carter, & Spuler, 1996), as well as by difficulties in academic and social adjustment (Nora & Cabrera, 1996). In her book *Teaching to Transgress,* bell hooks (1994) reflects on similar feelings of exclusion, disengagement, and depoliticization upon entering a desegregated school in the late 1950s—mirroring the psychological experiences of many students of color entering predominantly White institutions of higher education (Nagda, Kim, & Truelove, 2004; Brower & Ketterhagen, 2004). Thus a central problem facing higher education today is how to move from a status of desegregation, in which psychological effects threaten the success of underrepresented students, to a more integrated community, characterized by positive intergroup relations.

In response to the growing racial tensions on many campuses, a number of educational interventions have emerged that implicitly or explicitly promote positive intergroup relations. These programs represent myriad approaches, practices, and disciplines. A number of programs, for instance, work to enlighten students' awareness, understanding, and knowledge of different racial groups through exposure to multicultural perspectives found in literature, history, and the arts (Banks, 2001). Other interventions underscore the importance of intergroup contact as a vehicle to address and explore group differences, build cooperative communities, and recognize different perspectives and worldviews (Zúñiga, Nagda, & Sevig, 2002; Bruffee, 1999; Haugsby, 1991).

Despite the promise that these practices hold for improving intergroup relations on college campuses, there is no comprehensive understanding of how effective they are in reducing racial bias among student groups. Although numerous studies have examined the effects of interventions, no attempt has been made to synthesize the studies into an organizing framework that allows for a thorough investigation of overall quality and differential program effects. Many questions remain unanswered concerning who benefits from these educational interventions and whether student, institutional, or environmental factors influence their effectiveness. Given that many program interventions are now required of all students, understanding their effectiveness across various groups is of critical importance.

This review endeavors to answer the following research question: *How effective are the various types of educational interventions that are designed to reduce racial bias among students in higher education?* To respond to this question, the following subquestions will be addressed:

1. How is racial bias defined and measured?
2. Historically, how have researchers conceptualized the formation of racial bias?
3. What are the nature and quality of the research and evidence that link various educational interventions to racial bias?
4. How can research be improved to promote a greater understanding of the relationships between educational interventions and racial bias?

5. Given the insights gleaned from the above questions, what conceptual framework is appropriate for evaluating the influence of educational interventions on racial bias?
6. What questions remain for further exploration, based on gaps in the extant knowledge base?

Given the tentative future of many diversity initiatives (e.g., affirmative action policies), coupled with the current rise in racial tensions on many college campuses, there is an urgent need to understand which programs are most effective in reducing racial bias. Furthermore, in attempts to improve intergroup relations on campus, administrative decision making should be more empirically driven than assumptive based, although the quality of empirical evidence must be taken into consideration. Thus, by consolidating and making more accessible the research on various interventions, administrators will be in a better position to evaluate and meet their current and future commitments toward building more integrated campus environments. The significance of this review reflects the need to understand how campuses are currently achieving this goal and to determine what future efforts are needed to fulfill the original intentions of the *Brown* decision.

DEFINING AND MEASURING RACIAL BIAS

Defining Racial Bias

Racial bias can be more generally defined through the concept of intergroup bias. Hewstone, Rubin, and Willis (2002), for instance, refer to intergroup bias as the "systematic tendency to evaluate one's own membership group (the in-group) or its members more favorably than a nonmembership group (the out-group) or its members" (p. 576). Further, Dovidio et al. (2004) suggest that intergroup bias appears in different forms, ranging from attitudes and beliefs about another group to emotional reactions and behavioral dispositions toward particular members of a group or toward the group as a whole. Intergroup bias, therefore, incorporates four main components: prejudice, stereotypes, affective reactions, and discrimination (Dovidio et al., 2004; Mackie & Smith, 1998; Wilder & Simon, 2003). A closer examination of these components provides an important foundation in understanding the variant forms of racial bias found on many college campuses today.

Types of Bias

Prejudice

Prejudice typically is defined as a negative attitude, although theorists remain conflicted on the exact nature of the underlying characteristics that accompany such attitudes. Allport's (1954) interpretation of prejudice, for instance, is steeped in cognitive terms and refers to a negative attitude, based on faulty or inflexible generalizations, that is directed toward an individual or group. More contemporary theories (e.g., Stephan, 1985), however, are based on group membership, emphasizing that individuals make assessments and evaluations of others based on their attitudes or beliefs about the group to which the person belongs.

Traditionally, prejudice has been operationalized according to more general attitudinal models and includes a cognitive component (thoughts or beliefs about an attitude object), an affective component (feelings or emotion associated with

the attitude object), and a conative component (behavioral predisposition toward an attitude object; Esses, Haddock, & Zanna, 1993; Zanna & Rempel, 1988). Despite the complexity of this model, researchers such as Eagly and Chaiken (1998) posit that attitudes are formed and expressed on the basis of any one of these dimensions. Stephan and Stephan (2001) also suggest that although these three components are normally consistent with one another, they are not always congruent. An individual, for instance, may hold particular prejudiced attitudes or beliefs about an individual but behave in a nonprejudiced manner.

Stereotypes

Stereotypes represent a constellation of beliefs or characteristics about members of particular groups (Hamilton & Sherman, 1996; Hilton & von Hippel, 1996). Whereas many theorists conceptualize stereotypes as inaccurate beliefs resulting from irrational processes (Allport, 1954) or excessive rigidity and resistance to change (Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950), other theorists point to the functional quality of stereotypes in helping individuals to process information about people and simplify the complexity of their environment (Hamilton & Troiler, 1986). More recent conceptualizations point to the notion of "dispersion inaccuracy," which involves overattributing the extent to which a particular group shares a trait (Ryan, Park, & Judd, 1996). This phenomenon has been referred to as the "magnification of diversity" or the "outgroup homogeneity effect," emphasizing the tendency for individuals to perceive their own group as unique and heterogeneous while viewing outgroups as a homogeneous set of others (Mullen & Hu, 1989). Outgroups are most often perceived as homogeneous along more negative traits, and research has demonstrated that majority groups are more likely to display this bias than are minority groups (Oakes, Haslam, & Turner, 1994).

Stereotypes are typically activated automatically on the basis of phenotypical characteristics (e.g., race, age, and sex; Stephan & Stephan, 2001). Devine (1989) proposes that all individuals are exposed to their society's prevailing stereotypes during socialization and that, unless consciously overcome, those stereotypes are automatically activated in everyday life. As a result, information processing tends to be biased and strongest for well-developed stereotypes, such as those concerning race, religion, sex, and gender (Stangor & McMillan, 1992). In addition, Stangor and McMillan assert that stereotypes are maintained by the greater tendency for individuals to recall expectancy-confirming information about different racial groups, which affects how individuals notice, perceive, and store information. This bias is exacerbated by "illusory correlation"—a term coined by Hamilton and Rose (1980)—that is, a memory bias in which individuals overestimate the frequency with which certain groups (e.g., minority groups) engage in negative behaviors.

Affective Reactions

The affective dimension consists of more emotional or visceral reactions based on the social identity of a particular individual or group (Dovidio et al., 2004). Emotional reactions range from strong negative affect (Mackie, Devos, & Smith, 2001), to discomfort and anxiety (Stephan & Stephan, 1985), to lesser degrees of admiration and respect for groups different from one's own (Smith & Ho, 2002). Smith and Ho's research on attitudes toward Asian Americans demonstrated that even when societal stereotypes are positive (e.g., Asian Americans are intelligent,

industrious, and successful), White participants experienced a range of emotions depending on the perceived competitive threat (e.g., economic success) and overall appraisal of the targeted group. In addition, Stephan and Stephan's research demonstrated that high levels of anxiety are associated with avoidance behavior, increased stereotyping, and assumed levels of dissimilarity to out-group members. Finally, Wilder and Simon (2003) suggest that the valence, strength, and infusion of affect are important components to consider in understanding the influence of affect on intergroup relations. For instance, studies show that individuals in negative or positive, rather than neutral, mood states are more likely to engage in stereotyping, especially when interacting across race (Bodenhausen, Sheppard, & Kramer, 1994; Forgas & Moylan, 1991).

Discrimination

Discrimination, the final component of bias, involves an unjustified negative behavior toward members of a particular racial group. According to Allport (1954), discrimination occurs when individuals or groups of people are denied equality of treatment, despite their desire or wish for such equality. Fishbein and Azjen's (1975) theory of reasoned action suggests that discriminatory behavior is based on the interplay between attitudes (e.g., prejudice) and social norms. According to this theory, individuals weigh both the value of the intended outcome and the probability of success against their perceptions of the level of support they will receive from those most important to them. Other interpretations focus on discrimination as actions intended to preserve one's privileged position at the expense of a targeted group (Jones, 1972).

Measurement of Bias

Contemporary research suggests that the expression of bias occurs both unconsciously and intentionally (Blair, 2001; Devine, 1989). The relative strength and presence of different forms of racial bias, therefore, are measured through a series of instruments designed to tap into both explicit and implicit expressions of bias. Explicit measures of bias involve a conscious response by the participant (e.g., self-report response), whereas implicit measures are evaluations and beliefs that are automatically activated by the presence of stimuli associated with a particular attitude object (Dovidio, Kawakami, & Beach, 2003). Although intuitively one might assume that explicit and implicit measures are directly related, a meta-analytic research study by Dovidio et al. (2003) found only modest significant relationships ($r = .244$) among 27 studies.

Hewstone et al. (2002), in their review of major social psychological studies concerning intergroup bias, discovered that the majority of studies investigated mild forms of prejudice, using traditional self-report measures. Implicit measures, however, offer the promise of assessing the true extent of an individual's bias, which is critical given the social desirability bias inherent in most explicit measures. Hewstone et al. make the general argument that intergroup bias should be examined at three levels: public and personal (both explicit) and unconscious (implicit).

Measurements of prejudice typically include scales that isolate the cognitive, affective, and behavioral aspects that underlie prejudice (Dovidio, Brigham, Johnson, & Gaertner, 1996). Prejudice has been measured in both explicit and implicit ways.

Explicit measures often include thermometer readings that assess the likeability of particular racial groups (Mackie & Smith, 1998), as well as self-report items that measure one's comfort and openness toward diverse racial groups. Other measures tap into more covert forms of prejudice (e.g., the Modern Racism Scale, by McConahay, 1986) and include built-in measures to control for social desirability. In addition, some measures assess behavioral and nonverbal cues, as well as the presence of language bias (Maas, Castelli, & Arcuri, 2000). Implicit measures use more subtle assessments; for example, some instruments measure the electrical activity in facial muscles, which are known to have positive and negative associations (Dovidio & Fazio, 1992; Esses et al., 1993).

Stereotypes are typically measured using descriptive assessments of members of a group or through checklists that match specific characteristics or traits to different racial groups (see Katz & Braly, 1933, for the earliest version). Other more implicit tests use response-time measurements of unconscious associations between group labels or members and trait concepts (Dovidio et al., 2003). For instance, subjects are shown a representation of Black and White racial categories before making a decision about a positive or negative word. Faster response times are assumed to reflect a stronger association between a category and an evaluation.

Measuring affective reactions often involves simply asking respondents to indicate on a Likert-type scale (e.g., a 5- or 7-point scale) the extent to which they experienced a range of various types of emotions (Dovidio et al., 2004). More implicit measures have recently been used to assess more spontaneous forms of affect, such as heart rate and facial muscle activity (Vanman, Paul, Ito, & Miller, 1997).

Measurements of discrimination, according to Mackie and Smith (1998), have received little conceptual attention within the social psychological domain. The most commonly used behavioral measures involve social distance (e.g., seating distances), resource allocations (e.g., offers of aid), personal disclosure, competitiveness, and nonverbal behaviors (Dovidio & Gaertner, 1998; Dovidio et al., 1996).

HISTORICAL OVERVIEW OF THEORIES
EXPLAINING RACIAL BIAS FORMATION

Providing a portrait of the formation of racial bias situates the intervention literature within a theoretical context and offers further insight into the objectives and rationale for particular intervention strategies. In addition, the movement from early to more contemporary explanations of racial bias offers insight into what factors may account for the rise of bias-related incidents on college campuses. The majority of research on racial bias, however, has focused exclusively on the attitudinal variants of racial prejudice (cf. Hewstone et al., 2002), warranting a closer examination of those theories that highlight the formation of racial attitudes.

Early Conceptualizations

During the 1930s and 1940s the concept of racial bias was steeped in psychoanalytic theory. The prevailing belief during this era was that racial bias resulted from unconscious, intrapsychic conflict that was manifested through a variety of defense mechanisms (e.g., projection, displacement, and scapegoating; see Dollard, Doob, Miller, Mowrer, & Sears, 1939). Psychodynamic explanations eventually gave way to more individual-level theories rooted in personality traits such as authoritarianism (Adorno et al., 1950) and dogmatism (Rokeach, Smith, &

Evans, 1960). Major criticisms, however, were levied against these personality approaches, including their use of nonrepresentative sampling techniques, acquiescence bias (all items worded in an authoritarian direction), interviewer bias, and a bias toward studies of individuals of lower educational and socioeconomic levels (Brown, 1965).

In 1954, Gordon Allport published his seminal work *The Nature of Prejudice,* which offered an eclectic approach to the study of prejudice based on six distinct lenses: historical (broad social context); sociocultural (cultural and societal context); situational (immediate social forces); personal (personality dynamics originating in early childhood experiences); phenomenological (the influence of various forces on experiences, perceptions, and hypotheses); and stimulus object (explaining negative attitudes in terms of offensive qualities of the despised group). Allport's theory also provided a more complex understanding of the role of intergroup contact in reducing racial bias.

Societal Explanations

Allport's (1954) theory on the nature of prejudice shifted the study of racial bias away from individual-level theories, which could no longer account for the high levels of prejudice exhibited by individuals, groups, and social institutions in the South, to more sociocultural explanations, which pervaded the 1960s and 1970s. Two major theories dominated this period: realistic group conflict theory and social identity theory. These theories stress factors related to both increased competition and peer-group influence (cf. Dalton, 1991a).

Realistic group conflict theory posits that group conflict arises from competition for scarce resources (e.g., wealth or natural resources) and that a real or perceived threat initiates hostility toward the source of the threat (LeVine & Campbell, 1972). Although several studies support this theory (Sherif, Harvey, White, Hood, & Sherif, 1961), critics have identified a number of problems, including the presence of in-group favoritism in the absence of conflict and the difficulty of eliminating in-group bias even when doing so is in a group's best interest (Brown, 1995).

Social identity theorists challenge the group competition premise and contend that the inclination to discriminate is simply due to an individual's membership in a particular in-group (Sidanius, 1993). Through the process of social categorization (assigning traits to others based on group memberships) and self-esteem enhancement (favorably evaluating in-group traits and negatively evaluating out-group traits), individuals identify with particular social categories, giving rise to their social identity (Tajfel & Turner, 1986). Although this theory has gained considerable empirical support, several discrepancies remain, including frequent findings of out-group favoritism, difficulty in determining which traits will be discriminated against by the in-group, and a weak correlation between in-group identification and in-group favoritism (Brown, 1995).

Contemporary Approaches

Modern approaches to the study of racial bias recognize that despite evidence that the racial attitudes of White people are improving (e.g., Kluegel, 1990; Schuman, Steech, Bobo, & Krysan, 1997), racial bias is still present in more indirect and less overtly negative manifestations (e.g., Bobo, Kluegel, & Smith, 1997; Gaertner & Dovidio, 1986; Sidanius, 1993; Sears, 1988). Four contempo-

rary theories support these ideas: symbolic racism, aversive racism, laissez-faire racism, and social dominance. In particular, these theories relate to more institutionalized forms of racial bias, such as the perception of unfair treatment, as well as the fear of diversity and changing values of many of today's college students (cf. Dalton, 1991a).

Symbolic racism posits that racial bias results from negative affect toward a particular out-group in conjunction with the belief that members of the out-group violate traditional values held dear by members of the in-group (e.g., individualism or self-reliance; Sears, 1988). Opposition to race-based policies (e.g., affirmative action), for instance, is assumed to be based on a combination of these factors. Several criticisms of this theory have been levied against its conceptual and theoretical foundations, including the lack of a stable and consistent definition of symbolic racism among the theory's major proponents (Sidanius, Devereux, & Pratto, 1992). Furthermore, it is difficult to apply this theory in determining the causes for opposition to a particular social policy because of the theory's conflation of anti-Black affect and traditional American values, which have not been measured independently.

Aversive racism theory contends that individuals experience a contradiction between unacknowledged negative affect toward minority groups and an egalitarian, nonprejudiced self-image, which leads to anxiety and the avoidance of out-group members (Gaertner & Dovidio, 1986). This theory has been supported by a number of experiments (Gaertner & Dovidio, 1986) and has resulted in promising new scales designed to measures modern racism (e.g., the Modern Racism Scale, by McConahay, 1986).

According to laissez-faire racism theory, the decline of overt racism indicates not racial equality but, rather, the acceptance of institutionalized racism (e.g., resistance to policy efforts to improve racist conditions) under the guise of a free market (laissez-faire) economy (Bobo et al., 1997). Bobo and Kluegel (1997) tested this theory by using the 1990 General Social Survey and found ample evidence supporting it, including a tendency for younger generations to express less overt forms of racial bias (i.e., old-fashioned racism) than is expressed among older generations.

Social dominance theory conceptualizes racial bias as a product of belief systems that allow majority groups to "legitimize" their positions of power, privilege, and prestige (Sidanius, 1993). Research on this theory demonstrates that the extent to which individuals adhere to hierarchical conceptions of society is directly related to their levels of political conservatism, prejudice, and justification for the unequal distribution of social value (Sidanius, 1993; Sidanius & Pratto, 1999).

THE NATURE AND QUALITY OF EDUCATIONAL INTERVENTION RESEARCH

Criteria for Inclusion and Review of Studies

In determining the effectiveness of various educational interventions, a broad interdisciplinary search for research articles was conducted across multiple databases (e.g., ERIC, Educational Abstracts, PsycINFO). Key search terms were varied and included the following partial list of keywords and identifiers: *intergroup bias, racial bias, prejudice, stereotype, discrimination, diversity, multicultural, intervention, higher education, college,* and *students.* Additional studies were also obtained from source citations in relevant works. Studies were then selected if they

met three basic criteria: (a) They were empirically grounded; (b) they investigated an educational intervention and its influence on racial bias; and (c) they specifically focused on students in higher education.

To categorize the interventions, a content analysis was performed on each study, with a particular emphasis on the definitions and descriptions of various interventions. Four main categories emerged on the basis of this analysis:

- Multicultural course interventions
- Diversity workshop and training interventions
- Peer-facilitated interventions
- Service interventions

Table 1 depicts a typology of these categories and includes program characteristics found under each category. Interventions varied along several dimensions, including their explicit and implicit focus on racial bias. Multicultural courses and diversity workshops, for instance, are often explicitly designed to

TABLE 1
Typology of educational interventions

Program characteristics	Multicultural course intervention	Diversity workshop or training intervention	Peer-facilitated intervention	Service intervention
Exemplars	Required diversity courses Non-required diversity courses Ethnic studies Women's studies	Race/ethnicity workshops Diversity training	Peer-facilitated training Living–learning communities Intergroup dialogue Collaborative learning	Service-learning Community service Volunteer work
Racial bias goals	Explicit	Explicit	Explicit/ implicit	Implicit
Primary pedagogical focus	Content	Content	Contact	Contact
Educational processes	Didactic instruction Readings Written assign-ments Discussion Group exercises Role playing	Didactic instruction Role-playing Discussion	Experiential learning Group projects Group exercises Reflection Discussion	Experiential learning Reflection Discussion
Duration	Long term	Short term	Both short and long term	Both short and long term

reduce racial bias, whereas service interventions are less intentional in their programmatic design. Multicultural courses and diversity workshops often attempt to enlighten students through content-based knowledge, whereas peer and service interventions highlight intergroup contact as a primary pedagogical tool. Interventions also differed in terms of the educational processes used to achieve their course goals: multicultural courses and diversity workshops emphasized didactic instruction, whereas peer and service interventions focused more on experiential learning. Finally, the duration of many of the interventions varied, although both peer and service interventions include long- and short-term programming.

The review that follows is based on a simple taxonomy. First, interventions were categorized by program type and further divided on the basis of the characteristics provided in Table 1. Studies were then classified according to their methodological approach, which included quantitative, qualitative, and mixed-method studies. Within each of the methodological classifications, studies were generally organized according to their national or single-institution focus. Studies were then critically reviewed along several criteria: overall and differential effects, conceptual framework, research design and methodology, sample characteristics, measurement, and generalizability. The strengths and limitations of the various studies are highlighted throughout the review, and the overall and differential effects are summarized at the end of each section. The appendix also provides detailed information about the various measures of racial bias used in the reviewed studies.

Review of Higher Education Studies

Multicultural Course Interventions

The field of multicultural education embraces a plethora of different models for transforming students from a level of monocultural to multicultural awareness (Sleeter & Grant, 1999). Just as the models represent a diverse spectrum of theoretical perspectives, so too are the goals that underlie the varied approaches. There is some consensus, however, around the broad goals of multicultural courses: gaining knowledge of diverse groups, reducing prejudice and stereotypes, developing skills to work with diverse others, and challenging inequalities and injustices found in society (Banks, 2001; Bell & Griffin, 1997; Bennett, 1999).

This section explores four different multicultural course interventions based on their curricular location. Humphreys (1997) and Gaff (1991) suggest that there are two main options for how multicultural courses appear in the curriculum: as a required part of the curriculum or as non-required course in which diversity is infused into the curriculum. In addition, multicultural course interventions are found in "single-group" studies departments, such as women's studies and ethnic studies. Altogether, 22 studies were reviewed that investigated the effects of required and non-required diversity courses and ethnic and women's studies courses on students' racial bias.

Diversity Course Requirements

According to a recent national survey by Humphreys (2000), approximately 62% of today's colleges and universities have or are in the process of developing a diversity requirement for their undergraduate students. Yet, there remains a dearth of empirical evidence on the efficacy of these courses. In fact, only 7 stud-

ies (4 quantitative, 1 qualitative, and 2 mixed-method studies) were found that examined the effects of a diversity course requirement on students' level of racial bias. Of these, 2 studies reported positive effects (Chang, 2002; Palmer, 2000); 3 studies showed mixed results (Hathaway, 1999; Hasslen, 1993; Bidell, Lee, Bouchie, Ward, & Brass, 1994); and 2 studies found nonsignificant effects (Brehm, 1998; Henderson-King & Kaleta, 2000).

Quantitative studies. Of the two quantitative studies (Brehm, 1998; Henderson-King & Kaleta, 2000) that reported nonsignificant effects, Brehm implemented a modified Solomon four-group design to explain the impact of a diversity course requirement on White students' stereotypes of different minority groups. Using a convenience sample of 103 students from 12 courses, Brehm found no difference in pre-post stereotyping behavior both within and between the experimental and control groups. The within-group analysis of the experimental group showed that both women's studies and ethnic studies courses, as opposed to other variants of diversity course requirements, were most effective in attenuating negative stereotypes. Additional analyses revealed no difference among students with previous diversity course experience, although freshmen were less likely than seniors to endorse negative stereotypes.

Unlike Brehm (1998), Henderson-King and Kaleta (2000) used a nonequivalent control group design to explain the impact of a required diversity course on both White and non-White students' feelings and beliefs about different racial groups. Racial bias was measured along two separate scales: a feelings thermometer (Miller & Miller, 1977) that rates how respondents feel about a particular group (0 = cold, 100 = warm) and a measure of beliefs about racism (e.g., racism is still a problem in society). Students in the experimental group (diversity course) demonstrated no change in their feelings and showed a marginally significant ($p < .07$) increase in the strength of their beliefs about the existence of racism in society. The control group members (no diversity course taken), however, became significantly less favorable toward Latino/as and African Americans and showed no change in their beliefs about racism. Within-group analyses of the control group revealed that the negative change occurred only among White students. Unlike the Brehm study, this one showed no differences in class levels between the two racial bias measures.

One of the major advantages of Brehm's (1998) study was the use of a modified Solomon four-group design. In using such a design, Brehm was able to identify the presence of a testing bias, in that students in the posttest-only group were less tolerant in several areas than students in the pre-posttest group; this finding suggests that the questionnaire itself may have influenced students' tolerance levels, possibly softening or veiling the effects of taking a required diversity course. Although Henderson-King and Kaleta (2000) were not able to discern possible testing effects, they mailed surveys to students at home rather than administering the instruments in class. In this regard, the researchers were able to limit demand characteristics stemming from students' desire to please the instructor with their responses. Henderson-King and Kaleta also examined the effects of a diversity course across racial groups, although minority groups were aggregated into one all-encompassing group.

Of the two quantitative studies that found positive effects (Palmer, 2000; Chang, 2002), Palmer's study offered few methodological details (e.g., it did not dis-

close the measurements of racial bias used) to substantiate his findings. Chang's study, however, used an 8-item adaptation of the Modern Racism Scale (see McConahay, Hardee, & Batts, 1981) to examine the impact of a diversity course requirement on students' prejudicial attitudes. Using a cross-sectional design, a total of 15 courses that met the university criteria for a diversity course requirement were selected and randomly assigned as either pretest-only (beginning of term, $n = 7$) or posttest-only (end of term, $n = 8$). The results revealed a significant difference between the two groups: Students in the pretest group reported significantly higher levels of prejudice toward African Americans than those in the posttest group. This study, like Brehm's (1998) study, found that students with previous diversity course experiences showed no significant difference on the Modern Racism Scale, casting further doubt on the cumulative effects of diversity courses.

One of the strengths of Chang's (2002) study was the use of several covariates (i.e., race, gender, sex, and socioeconomic status [SES]) to control for possible interaction effects. Moreover, the randomization of the selected courses, with many pertaining to differences based on class, gender, or sexual orientation, gave credence to the relationship between learning about difference in general and a reduction in racial bias. The cross-sectional design used in the study, however, warrants caution in drawing any substantive conclusions. One reason is that such a design does not allow for a precise examination of student change across time. Finally, this study did not account for differences among racial groups, nor were measures of racial bias directed at groups other than African Americans.

All of the quantitative studies relied on convenience samples (based on instructors' and students' agreeing to participate) within a single institution, thereby limiting generalizability to other institutions. In addition, each study investigated a unique set of courses based on requirements particular to each institution, which further compromised the generalizability of the results. Finally, these studies offered little information on how factors such as the racial makeup of the class, the instructor's race, and the specific content or pedagogy employed may have affected the results.

Qualitative studies. Hathaway's (1999) dissertation used a phenomenological approach to investigate the impact of a required diversity course on nine White students' personal and societal beliefs regarding inequality. Results showed that the course helped students to reflect on and question the ways in which dominant thinking is socially constructed, although students demonstrated little change in their ability to reflect on or question their own personal belief systems regarding racial inequality. Although the analysis was quite rigorous and elaborate, the small number of students within a single institution limited the generalizability of the results. More problematic, however, are the threats posed to internal validity when a study relies on a single interpreter (e.g., the principal investigator) to substantiate the findings.

Mixed-method studies. Both of the mixed-method studies (Bidell et al., 1994; Hasslen, 1993) reported a mixture of positive and nonsignificant effects. Bidell et al. used a one-group pretest-posttest design to analyze White students' ($n = 55$) written responses regarding their understandings of the nature and causes of racism. The researchers couched the study in a conceptual framework that suggested that racial understanding followed a developmental sequence based on increasing levels of cognitive complexity. The results indicated a significant increase

in the complexity of students' understanding of racism, although in individual analyses only half of the students actually showed increases. Regression analyses showed that neither age nor class level were significant predictors of change scores (cf. Henderson-King & Kaleta, 2000); however, the small number of upperclassmen undermines this finding.

Hasslen's (1993) dissertation examined White students' ($n = 265$) responses to a cultural awareness inventory and situational attitude scale in conjunction with students' classroom writings and journals. Results showed an increase on 16 of the 28 cultural awareness measures, and 50% of the situational attitudes showed a significant difference. Hasslen further concluded, based on students' writings, that the increases were more likely based on increased sensitivity rather than actual reductions in prejudice. In particular, she found the presence of a history bias (i.e., the Los Angeles riots) that was associated with heightened sensitivity around racial issues.

Both studies looked exclusively at White students and relied on one-group pre-post designs. Although both included qualitative data, which aided in the overall interpretation of findings, the lack of control groups casts doubt on the causal nature of the course itself and increases the susceptibility to internal validity threats (i.e., selection bias). Thus, in both of these studies, it is difficult to discern if the effects are based on the course alone or if particular student characteristics or external factors are confounding the results. Like the quantitative studies, these studies relied on convenience samples and provided few details about the classroom environment, thus limiting their generalizability.

Non-Required Diversity Courses

Non-required diversity courses are similar to diversity requirements in terms of general goals and course objectives, although the non-required nature of these courses carries an additional burden in terms of controlling for selection effects. Of the 8 studies (5 quantitative, 2 qualitative, and 1 mixed-method) reviewed in this section, 5 reported positive effects (Inkeles, 1998; Marin, 2000; Khan, 1999; MacPhee, Kreutzer, & Fritz, 1994; Gurin, Dey, Hurtado, & Gurin, 2002); 1 reported nonsignificant effects (Taylor, 1994); and 2 reported a mixture of positive and nonsignificant effects (Lopez, 1993; Smith, 1993).

Quantitative studies. Five studies (Inkeles, 1998; Smith, 1993; Taylor, 1994; Lopez, 1993; Gurin et al., 2002) were examined that relied on longitudinal panel data from the same institution. The results, however, were quite discrepant. Both Inkeles and Smith, for instance, examined students' attitudes regarding affirmative action policies and practices. Inkeles looked specifically at Asian American students and found that they became more supportive of affirmative action after attending a class that included race and ethnicity issues in its curriculum. Smith, however, limited her study to White students and found a differential gender effect for women only. Taylor, on the other hand, looked at the impact of diversity courses on White students' level of tolerance and found nonsignificant results for both men and women. Finally, Lopez (1993) and Gurin et al. (2002) both examined the effects of a diversity-based course on White, Asian American, and African American students, although Lopez investigated students' awareness of racial inequality and Gurin et al. investigated racial engagement. Lopez found significant increases for White students only, whereas Gurin et al. found significant effects for all three racial groups.

The strengths of these studies rests in their use of longitudinal data as well as their inclusion of multiple student and institutional variables to control for confounding effects. These studies also suggest differential effects across racial groups, although the inconsistency across studies limits any definitive conclusions. One of the limitations across all of these studies is the independent variable that is used to measure course exposure. The item wording stresses courses that include discussions on race or ethnicity, although the face validity of this measure seems open to multiple interpretations. Finally, these results are limited in their generalizability to other research universities with similar institutional characteristics.

Qualitative studies. Both of the qualitative studies (Marin, 2000; Khan, 1999) reviewed reported positive findings. Marin used a case study method to understand how three courses (English, ethnic studies, and education) that infused a diversity perspective affected a range of educational outcomes, including the reduction of stereotypes. Using several qualitative techniques, Marin concluded that all three courses were effective in challenging and reducing racial stereotypes, developing critical thinking skills, and broadening student perspectives. Khan's study, however, relied on a more rudimentary analysis of students' final papers and course evaluations to draw conclusions about the efficacy of a diversity course. Her results showed that the course improved students' knowledge of the causes of racism and stereotyping as well as their understanding of how social norms contribute to racial bias.

Although these are encouraging results, the purposive sampling technique limits their generalizability to other courses. The generalizability is further hampered by the lack of descriptive information about the age, race, and gender of the samples used in both studies. Although the qualitative nature of these studies allowed for a deeper understanding of key course components (e.g., diversity content, active learning, and structured interactions) that were used to reduce racial bias, both failed to demonstrate how those course elements were linked to student learning.

Mixed-method studies. MacPhee et al. (1994) investigated the effects of incorporating a diversity perspective into a human development course on students' levels of old-fashioned and modern racism (see McConahay, 1986). Using a mixed-method research design, the researchers compared the effects of a diversity-based human development course to general courses in the social sciences, natural sciences, and business. In a primarily White sample, the researchers found significant differences in mean racial attitude scores, with students in the human development course showing less prejudice than the three control groups, even when controls were included for previous coursework and pretest scores. Class-level effects were also found (cf. Brehm, 1998), with seniors in the social sciences showing slightly higher levels of racial prejudice than 1st-year students, although the opposite trend was true for natural science and business students.

The scientific rigor of this study is impressive, especially the use of a mixed-method approach that included a pre-posttest design with three control groups, but the study has two noteworthy limitations. First, the results from White students and students of color were aggregated, which does not allow for intergroup comparisons. Second, because the number of covariates used in the study was limited (i.e., previous behavioral and social science courses), a selection effect may be confounding the strength of the course effect.

Ethnic Studies

A total of 7 quantitative studies that included an ethnic studies intervention were reviewed in this section. Of these, 5 reported positive findings (Astin, 1993a; Milem, 1994; Antony, 1993; Hyun, 1994; Hurtado, 2001), and 2 reported both positive and nonsignificant effects (Vogelgesang, 2001; Gurin et al., 2002). All of the studies reviewed in this section relied on data from the Cooperative Institutional Research Program (CIRP), which draws a large national sample of college students from a representative group of private and public 4-year colleges. In addition, each study measured the cumulative effects (frequency of taking ethnic studies courses) on students' racial bias, and in most cases, used Astin's (1991) college impact framework to assess student outcomes.

Quantitative studies. Five studies used CIRP data to explore the effects of ethnic studies courses on students' commitment to promoting racial understanding (Astin, 1993a; Milem, 1994; Antony, 1993; Hyun, 1994; Vogelgesang, 2001). Astin, Milem, Antony, and Hyun found positive relationships between ethnic studies courses and promoting racial understanding with several notable differences: Milem's study looked exclusively at how these courses affected White men and women and found significant, positive effects for both groups; Hyun examined effects separately for Whites and African Americans and found the same predictive value (β = .06) for both groups; and Antony and Astin found positive effects, although their results represent an aggregate of all students with no attention to racial differences. Among these studies, however, Vogelgesang's stands out. She examined changes in racial understanding across four different racial groups: Whites, Asian Americans, Latino/as, and African Americans. In taking a more unique, "critical" approach to quantitative analysis, she found that enrollment in an ethnic studies course was highly significant for White and Latino/a students, marginal for Asian Americans, and not significant for African Americans.

Both Hurtado (2001) and Gurin et al. (2002) used CIRP data to examine the relationship between courses in ethnic studies and students' awareness, appreciation, and acceptance of different racial groups. Using partial correlations to analyze the data, Hurtado found highly significant correlations between ethnic studies and all three student outcomes, although the correlations remained low (.11 to .19). Gurin et al. improved on Hurtado's study by including controls for the social environment (e.g., percentage of minority students and institutional emphasis on diversity) and selectivity as well as analyzing results by racial group. Results showed that enrollment in an ethnic studies course significantly influenced racial attitudes for White, Asian, and Latino/a students but had no effect on African Americans. When controlling for other diversity experiences, these results remained significant only for Whites and Asian Americans.

In general, the large representative samples used in all of these studies allow for some generalizability to other 4-year colleges, although two cautions are noteworthy: The CIRP data set is skewed toward private institutions; and low follow-up response rates (only 22% in Vogelgesang's study) suggest the strong likelihood of a sampling bias toward students who persist in college and participate in national studies. Several studies, however, included student background and institutional controls and statistical weighting techniques, which may have offset some of the problems associated with sampling bias. Many of these studies also relied on single-item self-report measures, which are subject to both measurement error and social

desirability biases. Finally, the large sample sizes found in CIRP studies and the small beta (regression) weights associated with many of the findings suggest a need to include effect sizes and to use more stringent criteria to indicate significance.

Women's Studies

Seven studies (6 quantitative, 1 mixed-method) were reviewed that investigated the effects of women's studies courses on aspects of students' racial bias. Of these, 4 reported positive results (Hurtado, 2001; Antony, 1993; Palmer, 2000; Astin, 1993a; Stake & Hoffman, 2001), and 2 reported nonsignificant effects (Vogelgesang, 2001; Hyun, 1994).

Quantitative studies. Like the research on ethnic studies, all of the quantitative studies used CIRP data to examine the cumulative effects of women's studies courses. Four studies (Hurtado, 2001; Antony, 1993; Palmer, 2000; Astin, 1993a) found that women's studies courses had a positive influence on students' racial bias, whereas 2 other studies (Vogelgesang, 2001; Hyun, 1994) found nonsignificant effects. The differences in results warrant a closer investigation into these studies.

Both Hyun (1994) and Antony (1993) used the same 1985–1989 CIRP data sets to examine the impact of a women's studies course on students' commitment to promoting racial understanding, yet Hyun found nonsignificant effects and Antony found significant, positive effects. A closer examination reveals several factors that may account for the different conclusions. First, Hyun used an initial sample of 24,878 students, but the actual results were based on approximately 10,000 students (presumably the difference reflects a listwise regression technique); Antony's sample and results were both based on 18,887 students, although no information was given regarding the treatment of missing data. Thus the difference in results may be partially due to the analytic sample. Second, Hyun used a more stringent indicator of significance ($p < .001$), whereas Antony reported results at a lower level of significance ($p < .01$). Finally, although both studies controlled for similar demographic variables (e.g., race, gender, and SES), Hyun's study incorporated a substantially higher number of controls, which may account for the differences in results.

Vogelgesang's (2001) study examined students' commitment to promoting racial understanding from a much more recent cohort of students (1993–1998) and analyzed the effects across racial groups. Therefore, race, cohort differences, or more contemporary women's studies courses may account for the lack of significant results. Hurtado's (2001) study, on the other hand, looked at cultural awareness, appreciation, and tolerance of others, which may have tapped more directly into outcomes of women's study courses. The studies by Astin (1993a) and Palmer (2000) were too descriptive in their presentation of results to make any substantive inferences.

Although many of the limitations found in these studies are similar to those carried out with ethnic studies, these studies emphasize two important considerations: First, using less stringent indicators of significance coupled with large sample sizes increases the chance of Type 1 error (alpha). Second, incorporating institutional and student background controls is essential in dampening potential confounding effects, and such decisions need to be grounded in findings from previous research studies.

Mixed-method studies. Stake and Hoffman (2001) conducted a classroom-based study looking at the effects of a women's studies course on students' egalitarian attitudes and awareness of discrimination (see Miller, Kinder, & Rosenstone,

1993). The researchers used a stratified sample of both women's studies (WS) and non–women's studies (NWS) students across 32 institutions (some private, some public, and varied in campus size from small to large), using quantitative and qualitative measures over three time periods. The study concluded that WS students expressed greater awareness of discrimination and more egalitarian attitudes than NWS across all three time periods. There were two other important findings: Critical thinking and open-mindedness were positively associated with both racial bias measures; and course content (over and above pedagogy, including participatory learning) was a major influence in student change.

The major strength of this study is the complexity of the research, which allows for some generalization to a variety of different institutional types as well as insight into the sustainability of course effects 6 months after completion. In addition, by including measures of teachers' characteristics and pedagogical areas, both, the researchers were able to determine that although pedagogy plays a role in changes in racial bias, course content plays a more significant role. Several limitations are noteworthy, however. First, the study did not control for previous diversity course experience or student background characteristics. Second, WS courses were not chosen on the basis of any criteria offered by the researchers, but rather by relying on localized definitions of WS courses at various institutions. Finally, the researchers neglected to look at differential effects across race or gender, calling into question the efficacy of the courses for change across non-White racial groups.

Overall Effectiveness of Multicultural Interventions

As a whole, the majority of studies support the conclusion that multicultural interventions are effective in the context of the higher education curriculum. Most of the survey-based research studies, which carry the highest ability to generalize to other institutions, support the cumulative effect of ethnic and women's studies interventions, although the results are less clear for diversity-infused courses. In contrast, the quasi-experimental, classroom-based studies show mixed results for diversity course requirements, which are the least susceptible to selection bias; the results of these studies also cast doubt on the cumulative impact found in the national studies. What remains unclear across these studies, however, is the effectiveness of multicultural interventions across race, gender, disciplinary, and class-level groups; discrepant findings across all of these factors were found in numerous studies. Whereas both qualitative and mixed-method studies add dimensionality to the understanding of effectiveness, especially with regard to pedagogy and developmental processes, they add little evidentiary weight to the differential race, class level, and gender effects of these interventions. Given the incipient nature of the research, these studies seem to raise more questions than definitive answers regarding the effectiveness of multicultural interventions.

Diversity Workshop and Training Interventions

Diversity workshops represent another important educational intervention currently used by colleges and universities to improve intergroup relations on their campuses. Diversity workshops differ from more traditional academic modes of instruction in several ways: They are more interactive, typically are short-term (2 hours long, on average), and use participants' own experiences with bias or discrimination as an important pedagogical tool (McCauley, Wright, & Harris, 2000).

Despite being described as "arguably the fastest growing innovation in the history of U.S. higher education" (McCauley et al., 2000, p. 113), empirical studies evaluating these programs remain at an incipient stage of development. Of the 11 studies (10 quantitative, 1 mixed-method) reviewed in this section, 9 reported positive effects (Pascarella, Edison, Nora, Hagedorn, & Terenzini, 1996; Whitt, Edison, Pascarella, Terenzini, & Nora, 2001; Springer, Palmer, Terenzini, Pascarella, & Nora, 1996; Milem, 1994; Antony, 1993; Hyun, 1994; Gurin et al., 2002; Astin, 1993a; Katz & Ivey, 1977), 1 reported nonsignificant results (Neville & Furlong, 1994), and 1 reported both positive and nonsignificant results (Vogelgesang, 2001).

Quantitative Studies

Of the 8 quantitative studies that reported positive effects (Pascarella et al., 1996; Whitt et al., 2001; Springer et al., 1996; Milem, 1994; Antony, 1993; Hyun, 1994; Gurin et al., 2002; Astin, 1993a), each relied on national databases (both CIRP and the National Study of Student Learning) and several investigated whether these effects remained consistent across gender and race. Milem, for instance, found that workshop participation predicted stronger effects for women in terms of their commitment to promoting racial understanding, whereas Hyun found stronger effects for White students than for African Americans. Similarly, Pascarella et al. and Whitt et al. found that workshop participation in the 1st year was associated with larger gains in openness to diversity for White students than for students of color, although these conditional effects dropped out in the 2nd and 3rd years. Vogelgesang (2001), however, found highly significant effects for Asian Americans and White students ($p < .001$), significant effects for African Americans ($p < .01$), and nonsignificant effects for Latino/a students.

Although these national studies attest to the cumulative benefits of workshop participation (i.e., frequency of participation), Whitt et al. (2001) demonstrated that even when controlling for students' workshop participation in the 1st or 2nd year of college, participation in the 3rd year had a positive, significant effect on students' openness to diversity. Springer et al. (1996) also found that although certain factors were more predictive of the likelihood of attending a workshop (i.e., liberal majors over conservative ones, higher SES, and higher degree aspirations), there were no interaction effects among these factors and workshop participation that influenced students' racial attitudes.

The results from the national studies carry limitations similar to those previously discussed: sampling and mortality biases, confounding effects based on differential controls, measurement error, and social desirability biases. Moreover, these studies offer little insight into what aspects of the workshop experience fostered student change and little information on the general nature of the workshop experience (e.g., voluntary versus mandatory). Most of these studies, however, used large, representative samples, which allow for some generalizability across institutional types.

Unlike the survey-based studies, Neville and Furlong (1994) used a quasi-experimental, posttest-only research design to examine racial attitude change across five racial groups of 1st-year students. The researchers employed two attitudinal measures of racial bias (i.e., the Social Scale and a modified version of the Social Scenarios Scale; see Byrnes & Kiger, 1988) and found no significant differences among any students in the experimental or control groups. Although no gender differences were found, African Americans in both the treatment and control groups

were more willing than Whites and Asian Americans to confront racially insensitive behaviors in social settings. Although this study incorporated a sound research design with two reliable and valid instruments, student attrition was extremely high for the experimental group and the results are likely reflective of this mortality bias. Moreover, mean scores on the racial bias measures were high for all groups, suggesting that a social desirability bias may create a ceiling effect that masks the true effects of workshop participation. Finally, a major drawback of a posttest-only study is the lack of control for selection effects, which suggests that other sample characteristics may be confounding the true effects.

Mixed-Method Studies

Katz and Ivey (1977) conducted one of the pioneering studies that examined how a racial awareness workshop influenced White students' racial attitudes and behaviors. Results showed that students who went through the training had more positive racial attitudes (based on the Steckler Anti-Black and Anti-White Inventory [Steckler, 1957] and the Attitude Exploration Survey [Adams, 1973]) and behaviors when compared with the control group, as well as a higher level of racial consciousness and awareness (the latter was based on content analysis of students' journals). Furthermore, a 1-year follow-up showed that these changes had been maintained. Although these results are encouraging, there are several notable limitations. Students' background characteristics (especially gender), for instance, were not listed or controlled for in the study, hampering the replication and generalizability of the results. Furthermore, how students were selected for the study (e.g., voluntary, mandatory, or for credit), as well as how specific aspects of the training influenced the results, were not explained or examined in the study.

Overall Effectiveness of Diversity Workshop and Training Interventions

The majority of studies examined in this section attest to a cumulative effect of diversity workshops and training on students' racial bias. However, even among studies that reported positive effects across racial groups, the strengths of the effects differed by racial group. Across studies White students tend to be most affected by workshop participation, although the differential effects across other racial groups are difficult to entangle because of the tendency for researchers to aggregate minority students into one group. As a whole, these studies suggest greater benefits for White and Asian American students and smaller or nonsignificant effects for African Americans and Latino/as. Only one study examined gender effects, but women also seemed to benefit more from workshop interventions. Despite the lack of interaction effects found for academic major in one study, the overall lack of attention in these studies to either class level or discipline makes it difficult to draw any definitive conclusions. It should be noted, however, that although the majority of these studies underscore the predictive nature of diversity workshops, only two studies employed quasi-experimental designs to explain classroom-based change. Although the two studies came to different conclusions, the study by Neville and Furlong (1994) carries more weight based on the contemporary nature of the findings. Thus the evidence points to the predictive power of diversity workshops but falls short in explaining the efficacy of these interventions to impart actual classroom-based change in students' bias.

491

Peer-Facilitated Interventions

Much has been written concerning the potential impact of various peer interventions on students' racial bias (Dalton, 1991b). In addition, several empirical studies have examined more broadly the benefits of cross-racial interaction on students' commitment to promoting racial understanding (Hyun, 1994; Milem, 1991) and openness to diverse perspectives (Pascarella et al., 1996; Whitt et al., 2001). Other studies have found differential effects based on both the quality of interaction (Smith, 1993) and students' racial group membership (Lopez, 1993). In this section, four peer interventions are explored: peer-facilitated training, living-learning communities, intergroup dialogue, and collaborative learning.

Peer-Facilitated Training

Peer-facilitated training and instruction involve the use of peers to train and/or instruct the subjects on issues of racial bias. The methods and theory underlying such approaches are quite varied, but most are based on sharing and discussing personal stories in conjunction with role-playing or other experientially based activities (Brown & Mazza, 1991). These programs differ from both diversity workshops, which are often facilitated by faculty or student affairs staff, and intergroup dialogues, which are more structured and sustained over a longer time period. Although much of the evidence for the effectiveness of peer-facilitated training is anecdotal (e.g., Vohra, Rodolfia, De La Cruz, Vincent, & Bee-Gates, 1991), one quantitative study (Nelson, Johnson, Boyd, & Scott, 1994) and one qualitative study (Pence & Fields, 1999) were reviewed; both reported positive findings.

Quantitative study. A study conducted by Nelson et al. (1994) investigated the effects of a short-term (1–3 hours), peer-facilitated diversity training on a group of undergraduate White students. Using a 2 × 2 research design, the results showed that after receiving the training, students in the experimental group were more optimistic about intergroup understanding, more comfortable interacting with minority students, and less likely to perceive minority students as unqualified to be at the university; the control group's attitudes became significantly more negative over time. Although the program was largely successful, the study has several limitations that warrant caution in interpreting these results. First, a 2 × 2 research design does not identify change in individual students; rather, it relies on two distinct pretest and posttest groups to measure change. Second, no equivalency controls (e.g., students' background characteristics) were used in the study, which suggests that other factors may be confounding the results. Finally, the variant time lag (1–4 weeks) in which posttests were administered increases the risk of a local history bias.

Qualitative study. Pence and Fields (1999) conducted a two-part experiment in which senior-level students conducted fieldwork on variant forms of institutionalized discrimination. The fieldwork involved a group of students observing the differential treatment a White and African American student received from a department store, car dealership, and apartment manager. Based on the field experience, which turned up both subtle and overt forms of institutionalized discrimination, the students presented their findings and reactions to a group of introductory 1st- and 2nd-year sociology students. Although the scientific rigor of the study is minimal at best, the results indicated several beneficial effects: Students in the introductory class were able to acknowledge aspects of discrimination and prejudice in their own communities, while questioning their own assumptions of an American meritocracy.

492

Living-Learning Communities

Living-learning communities (LLC) differ from traditional residence halls in that the former interweave educational opportunities and scholarship into the students' living arrangements (Blimling, 1993; Pascarella, Terenzini, & Blimling, 1994). LLCs provide extended opportunities for students (and faculty) to interact (Romanoff, 2000), as well as exposure to programming and planning around multicultural issues (Hughes, 1994). Johnson and Johnson (2000) contend that prejudice and discrimination are optimally reduced when students participate in cooperative communities, although there is a dearth of empirical research backing such claims. Earlier reviews have been conducted on LLCs (see Blimling, 1993, for a more general review), but linkages between LLCs and racial bias were only minimally explored. In this section, three quantitative articles (Pascarella & Terenzini, 1980, 1981; Pike, 2002) are critically reviewed, all of which reported positive outcomes.

Quantitative studies. Two quasi-experimental studies by Pascarella & Terenzini (1980, 1981) compared the impact of an LLC to that of more traditional residential programs for students across several academic, personal, and behavioral measures. Using single-institution data from a representative group of students (range of 536–763 between two studies), the 1980 study revealed that the LLC student group had a significant positive association with the personal development scale, which included items such as students' interest in and openness to new ideas, their self-awareness, and their interpersonal skills; nonsignificant findings, however, were uncovered in the 1981 study. The 1980 study also found that when student-to-student and faculty-to-student interaction measures were added to the analysis, the effects of the LLC program on students' personal development became insignificant. Thus the researchers concluded that the structural and organizational influence of the LLC is mediated by both the quality and the impact of students' interactions with faculty and peers.

Pike's (2002) study examined how LLCs influence students' openness to diversity. Using a sample of 502 1st-year undergraduate students, the study used a posttest-only control-group design to uncover the effects of two types of LLC on students' openness to diversity. The results showed that LLC participation produced a stronger overall effect on student's openness to diversity in comparison with traditional residence halls; a significant part of the effect was mediated by students' high levels of positive peer interaction. It is important to note, however, that compared with off-campus housing, both LLCs and traditional residence halls were significant, positive predictors of students' openness to diversity. The results, however, are based on a cross-sectional design that does not shed light on the actual nature of causality. Given the sampling bias (only a 30% return rate), selection bias (students self-selected different residence types), and single-institution focus, generalizations about the results are most likely limited to larger research universities with comparable residential populations.

Intergroup Dialogue

Intergroup dialogue programs bring students from diverse social identity groups (e.g., race, gender, sexual orientation) together to promote communication across difference, to practice constructive intergroup relations and coalition building, and to develop skills necessary for working and living in multicultural communities (Schoem, Hurtado, Sevig, Chesler, & Sumida, 2001; Vasques Scalera, 1999; Zúñiga & Nagda, 1993). Intergroup dialogues differ from more traditional forms

of multicultural education based on "feel good" approaches (cf. Sleeter & Grant, 1999) by foregrounding intergroup conflict and difference (Vasques Scalera, 1999). Dialogue programs rely on peer-facilitated, face-to-face meetings of students from different identity groups to explore group differences, challenge stereotypes and misinformation, and address issues of intergroup conflict (Zúñiga et al., 2002).

Of the 12 studies examining intergroup dialogues (8 quantitative, 3 qualitative, and 1 mixed-method), 7 studies reported positive findings (Gurin, Peng, Lopez, & Nagda, 1999; Lopez, Gurin, & Nagda, 1998; Nagda et al., 2004; Nagda, Gurin, & Lopez, 2003; Gurin, Nagda, & Lopez, 2004; Geranios, 1997; Vasques Scalera, 1999; Nagda, Spearmon, & Holley, 1999), 3 studies reported mixed results (Spencer & Nagda, 2002; Yeakley, 1998; Alimo, Kelly, & Clark, 2002); and 1 study reported nonsignificant findings (Nagda & Zúñiga, 2003).

Quantitative studies. Of the 8 quantitative studies, 4 relied on freshman data from a similar dialogue course at the University of Michigan (Gurin et al., 1999; Lopez et al., 1998; Nagda et al., 2003; Gurin et al., 2004). With the exception of the Nagda et al. study, these studies employed nonequivalent, matched control-group research designs that included primarily White, female, 1st-year students. Using several different measures, researchers found consistent effects for both White students and students of color who participated in dialogues: greater commonality and less divisiveness among different groups, heightened racial awareness, more support for affirmative action and multicultural programs, and increased awareness of the structural causes of inequality. Gurin et al. (1999) and Gurin et al. (2004) also investigated the long-term effects of these changes and found that students who participated in dialogues were still significantly higher than the control group along several outcomes (e.g., racial awareness and appreciation of group differences and commonalities) 4 years later.

In addition to the University of Michigan studies, 3 studies relied on data collected from both undergraduate and graduate social work students (Spencer & Nagda, 2002; Nagda et al., 2004; Nagda & Zúñiga, 2003). Spencer and Nagda looked at a group of primarily White undergraduate and graduate women and discovered a differential class-level effect: Of the eight race-based measures, undergraduates showed significant increases on all but one measure (perspective-taking), whereas the graduate students showed significant increases along only two measures. The researchers offered three rationales for the differences: Graduate students entered with higher mean scores (ceiling effect); the graduate sample was significantly smaller; and the graduate course emphasized multiple identity groups, whereas the undergraduate course focused on race. In addition, given the differences in course instructors, institutions, and course length, other factors may be confounding the class-level effect.

The 2 other social work studies are important because their conclusions were quite different. Nagda and Zúñiga (2003) examined a group of 42 undergraduate students (81% women, 55% students of color) and found nonsignificant course effects overall, although they believed this was due to a ceiling effect (especially among students of color, who scored very high on measures of racial identity). Nagda et al. (2004) looked at a much larger sample ($n = 231$) of undergraduate students (85% women, 39% students of color) and found significant course effects for all students. The different findings are likely due to several methodological differences: Sample sizes differed in both studies; course and

instructor characteristics varied across studies; and the outcomes also were different (one focused more on the centrality of race whereas the other looked at prejudice reduction more specifically). Both studies, however, failed to incorporate control groups, which opened them to selection biases and limited the causal nature of the findings.

All of the quantitative studies relied on self-report measures to explain change, an approach that may result in ceiling effects that distort true classroom-based change. Moreover, many of the studies are susceptible to testing biases based on the short duration of intergroup dialogue programs. Despite these limitations, many of them offer insight into the efficacy of different pedagogical processes (content-based pedagogies and the dialogic process) in eliciting student change. And, despite the tendency to aggregate students of color, many of these studies examined differential effects among White students and students of color.

Qualitative studies. Of the 3 qualitative studies, 2o were dissertations (Vasques Scalera, 1999; Yeakley, 1998) that used a grounded theory approach. Vasques Scalera examined the impact of a dialogue program on a group of 30 student facilitators (19 women, 12 students of color) 1 to 4 years after the course. The results demonstrated that the biggest changes for students related to an increased recognition of their own biases, stereotypes, and misinformation. In particular, White students learned more about the structural causes of oppression, whereas students of color gained insight into how their own experiences with oppression affect how they interact with White students. Yeakley, however, examined a group of 13 students (7 females, 7 students of color) enrolled in four different dialogue courses. Her results yielded both positive (e.g., increased intergroup understanding) and negative (e.g., increased stereotyping and separation) themes, although she postulated that underlying these changes was the degree of social distance that students experienced with diverse others.

Together, the qualitative studies seem to further substantiate the cognitive and affective gains found in the quantitative studies, although both studies relied on retrospective self-reporting and single sources of data to substantiate their findings. Yeakley's study is particularly valuable in that she reported on both positive and negative findings. Given the time lag in Vasques Scalera's study, a more in-depth probe of rival explanations (e.g., other experiences, maturation, or history) would have further substantiated her findings. Finally, the small and purposive sample sizes used in these studies limits their generalizability.

Mixed-method studies. Only 1 study (Nagda et al., 1999) incorporated a mixed-method design to examine the impact of an intergroup dialogue intervention. Researchers used a mixture of surveys, focus groups, and in-depth interviews to determine how a dialogue course affected a group of 50 juniors (84% women, 50% White) in social work. As this was an exploratory study, little detail was included regarding the methodology used to analyze the data, and results were limited to posttest descriptions of the top five learning goals. More than 93% of the students stated that they had gained knowledge of diverse groups, learned to value diverse viewpoints, increased their awareness of social inequalities, and understood the impact of social group membership. Qualitative reports also emphasized the importance of the dialogic process in enhancing learning, as well as the motivation and preparation of the peer facilitators. Racial differences, however, were not discussed and little evidence was presented to substantiate the learning themes.

Collaborative Learning

Although collaborative learning has been studied extensively in both elementary and secondary education (Slavin, 1995), it has only recently emerged in higher education as a promising educational practice (Cockrell, Caplow, & Donaldson, 2000). Collaborative learning emphasizes small-group work over traditional lecture formats and requires intensive interaction among students as they work through complex projects and exercises (Bruffee, 1999). Several studies have shown the impact of collaborative learning on student learning and achievement (e.g., Tinto, 1997; Astin, 1993b), but only recently have arguments been made about the links between collaborative learning and stereotype reduction in higher education (Cabrera et al., 2002; see also Vogt, 1997, for tolerance link). In addition, while numerous studies advocate the benefits of collaborative learning (e.g., Shapiro & Levine, 1999), empirical proof is only beginning to surface. Therefore, only 4 research studies (3 quantitative, 1 qualitative) were included in this review, of which 2 reported positive results (Cabrera et al., 2002; Cabrera, Nora, Bernal, Terenzini, & Pascarella, 1998) and 2 reported mixed results (Wolford & Clemo, 1997; Lawrence, 1998).

Quantitative studies. Of the 3 quantitative studies reviewed, 2 were based on multi-institutional data from 23 private and public colleges (Cabrera et al., 2002; Cabrera et al., 1998). Data for these studies were drawn from the National Study of Student Learning, which included a predominantly female (64.5%) and White (62.2%) group of 2,050 students in their 2nd year of college. Researchers used two different measures of collaborative learning, based on students' preferences for and experiences with different collaborative learning practices, and one multi-item measure to assess students' openness to diversity. Results from both studies revealed two important findings: There were no race or gender differences in students' preferences for collaborative learning; and, controlling for a number of background and different learning outcomes, collaborative learning had the largest and most significant impact on students' openness to diversity. Although the limited number of institutions prevents wide generalizability, these studies make a strong case for the cumulative benefits of collaborative learning experiences.

The other quantitative study by Wolford and Clemo (1997) examined the impact of a classroom-based collaborative project on students' racial bias. Comparing the results from the same course in two different years (1993, when it was an elective, and 1997, when it was required), the researchers found conflicting results: The 1993 sample showed significant pre-post change in prejudice reduction, whereas the 1997 sample showed nonsignificant change. The samples were relatively similar (predominantly White and female), but the 1997 sample was younger, which suggests that a maturation bias may account for some of the discrepancy in results. Moreover, the nature of a required versus an elective course suggests that the latter may be more prone to selection biases. Finally, the researchers suggest that a concomitant co-curricular diversity event series during the 1993 course may have also confounded the results.

The results of the quantitative studies yield mixed conclusions about the efficacy of collaborative interventions. Although the multi-institutional studies provide strong evidence of a cumulative effect, the model under investigation explained only 9% of the variance in students' openness to diversity. In addition, the findings offer little insight into what factors (e.g., race/gender makeup of groups, type of collaborative undertaking, or instructor characteristics) mediate the efficacy of var-

ious collaborative learning experiences. Wolford and Clemo's (1997) study is important in that it underscores the necessity of controlling for selection, maturation, and history biases. Without such controls, the quality of the research and evidence are seriously undermined, casting doubts on the overall efficacy of a particular intervention.

Qualitative studies. The 1 qualitative study, Lawrence (1998), investigated the effects of a collaborative learning exercise on students' racial bias. Lawrence had a group of 21 students (White, female, upperclass level) participate in a collaborative exercise in which they were given differential supplies (some lavish supplies, some very minimal supplies) to create a mobile that expressed the meaning of multiculturalism. Students worked in teams to create their mobiles and then presented them to the class. The qualitative results, based on journal reflections and class observations, were mixed: For 15 of the 21 students, the experience profoundly affected their views of themselves and others and increased their awareness and understanding of their privileged positions in society. This study sheds some light on collaborative interventions; however, the methodology used to interpret the results was not discussed, nor was compelling or ample evidence provided to back up the instructor's claims.

Overall Effectiveness of Peer Interventions

In general, the evidence points in the positive direction for the effectiveness of peer interventions to reduce bias. With the exception of studies of intergroup dialogue, the small number of studies uncovered in other areas warrants caution in drawing any definitive conclusions. The larger number of intergroup dialogue studies seems to support their general efficacy across racial groups, although that efficacy may be limited to undergraduate students. Little is known about the efficacy of other peer interventions across minority groups, and because most samples favored females, these studies offer little insight into differential gender effects. As in other sections in this article, survey-based studies generally support the cumulative effects of peer interventions, but the quasi-experimental studies show mixed results in terms of student change. The designs of several of the studies also suggest that a number of internal and external threats to validity (e.g., selection, social desirability, history) may be distorting and confounding the true effects of these interventions. Although the qualitative and mixed-method studies, especially those on intergroup dialogue, provide a deeper understanding of which aspects of racial bias are positively and negatively influenced by program interventions, many are prone to methodological weaknesses. Thus peer interventions seem to offer the potential to reduce racial bias, but the incipient nature of the research contributes to only a broad understanding of the effectiveness of these programs.

Service Interventions

Although the goals of many service interventions are not explicitly directed toward racial bias, many programs help students to overcome bias through direct contact with those who have been defined as the "other" by the dominant culture (Haugsby, 1991). Etzioni (1983) argued that service programs provide opportunities for individuals from diverse racial and ethnic groups "to get to know one another on an equal footing while working together at a common task" (p. 161). Service-learning programs, in particular, challenge students to engage actively in

dialogues about issues of "equity, difference, inclusion, tolerance, justice, and power" (Saltmarsh & Heffernan, 2000, p. 5).

A total of 13 studies (6 quantitative, 2 qualitative, and 5 mixed-method) were reviewed to determine the effectiveness of service programs in reducing students' racial bias. With the exception of 1 study that reported mixed results (Dunlap, 1998), all of the studies support the general conclusion that participation in service is an effective means of reducing racial bias.

Quantitative Studies

The studies by Astin and Sax (1998) and Astin, Sax, and Avalos (1999) both used CIRP data to determine the impact of service experiences on students' openness to diversity, although there are notable differences between the two studies. Astin and Sax, for instance, found that four types of service experiences (i.e., education, human needs, public safety, and the environment) were significant in promoting racial understanding, increasing knowledge of different racial groups, and enhancing students' ability to relate to people of different races and cultures. The study by Astin et al., however, investigated the long-term effects of service participation and found that even after controlling for a number of background and experiential variables, students who participated in service in college were more likely than nonparticipants to promote racial understanding and socialize with diverse others 5 years after college.

The 2 other multi-institutional studies, by Eyler, Giles, and Braxton (1997a, 1997b), investigated a group of 1,500 students at 20 universities. In the first study, the researchers examined both the entire group and a subset of 404 students with concentrations in the liberal arts. Using a number of control variables, the researchers found that service experiences were a highly positive predictor of posttest tolerance scores (see Walt Whitman Center for the Culture and Politics of Democracy, 1993, for details on measure) for both the entire sample and the liberal arts subsample. A selection effect was also discovered, in that service participants scored significantly higher on the pretest than did nonparticipants across a number of dimensions, including attitudes, skills, values, and understanding. The second study also found that the quality of the placement (e.g., level of challenge) was a significant determinant of students' posttest tolerance toward diverse others.

The results from these 4 studies point to the cumulative effects of service programs across a number of types of institutions, but several limitations permeate all of the studies. First, the beta coefficients studies were very small, suggesting a negligible overall effect for service participation. Second, the studies relied on single-item measures for their dependent variables, which carries an increased risk of measurement error. Third, differences across racial groups were not explored. Finally, most of these studies aggregated distinct service programs (e.g., volunteerism, community service, and service-learning) into one service category, an aggregation that leaves many questions unanswered about the differential effects of program types.

Of the 2 classroom-based studies, the study by Everett (1998), was limited to a cursory analysis of course evaluations, although the vast majority of students cited learning more about their personal values and beliefs surrounding social inequality. Myers-Lipton (1996), however, performed one of the most rigorous and methodologically sound studies, on the impact of service-learning on students' levels of modern racism (using the Modern Racism Scale). Using a nonequivalent control

group design, Myers-Lipton compared three groups of students: those who were involved in a service-learning program (SL), those who were involved in a service program without a learning component (SNL), and those who were not involved in any service program (NS). The results showed that students in the SL groups decreased their pre-post levels of modern racism and showed significant decreases in comparison to the SNL and NL groups, although differences between the NL groups were strongest. Post-posttest results showed similar trends 2 years later.

Several methodological strengths of the Myers-Lipton (1996) study are noteworthy. First, to control for internal threats to validity, the dependent variable was regressed on three separate sets of independent variables: The first set included only the experimental and control groups; the second set included pertinent background variables (e.g., gender, race, political orientation) in addition to the student groups; and the third set included the pretest variable in addition to both background and student group variables. In this way, the researcher minimized internal validity threats due to both self-selection and sampling biases. In addition, the use of a valid and reliable measure of racial bias, such as the Modern Racism Scale (see McConahay, 1986), provided greater insight into the efficacy of service interventions. Despite the scientific rigor of the study, however, rival explanations for the findings were not adequately ruled out. For instance, it may be that the SL groups were involved in specific service experiences that included substantial interactions with different racial groups, whereas the SNL group may have had limited exposure to different racial groups.

Qualitative studies. Of the 2 qualitative studies (Rhoads, 1998; Dunlap, 1998) reviewed, Rhoads used a phenomenological methodology to examine the type of learning that occurred for students (85% White, 80% female) across three institutions. Using a blend of formal and informal interviews, open-ended surveys, participant observations, and documents, Rhoads uncovered three general themes in students' experience of service work with individuals of diverse racial and cultural backgrounds: confrontation of generalizations, eradication of stereotypes, and reexamination of preexisting prejudicial attitudes. Rhoads's study, however, does not infer causality (other factors may also have contributed to students' learning), nor does it differentiate service-learning from other service-based experiences or look at experiences across student groups. The credibility of the study, however, is enhanced by the use of multiple sources of data and member-checking to triangulate and substantiate the findings.

Dunlap's (1998) study used students' reflection journals to determine the impact of a service-learning course. Based on a group of 30 undergraduate students (25 White, 24 female), Dunlap found that more than 70% of the students' journals contained references to multicultural issues and 30% of the entries reflected on a particular racial or cultural incident, which was often accompanied by an emerging awareness of and sensitivity about societal oppression. Although minority students' journals showed little evidence of an emerging awareness or questioning of their racial attitudes and beliefs, the reasons behind this omission were not explored. In addition, the study relied solely on students' journals, which increased the study's susceptibility to demand characteristics. Overall, the studies by Rhoads (1998) and Dunlap suggest the importance of reflection and interaction (both integral parts of service-learning) in raising students' awareness and understanding of issues of power, privilege, and societal oppression.

Mixed-method studies. In addition to the above studies, 5 mixed-method studies were also reviewed. Of these, 1 was multi-institutional (Eyler & Giles, 1999); 1 presented preliminary results from a larger study focused on a campuswide assessment plan for service-learning (Driscoll, Holland, Gelmon, & Kerrigan, 1996); and 3 included a mixture of surveys and open-ended questions to assess students' learning (Giles & Eyler, 1994; Rice & Brown, 1998; Osborne, Hammerich, & Hensley, 1998). All of the studies reported positive results, and although several studies included significant proportions of minority students (range of 17–57%), none reported on differences across racial groups.

Eyler and Giles's study (1999) examined more than 1,600 students (68% female, 83% White) at 20 institutions, using both surveys and structured interviews. The results, based on more than 100 different service-learning courses, revealed that the most consistent student outcomes were reductions of negative stereotypes and increases in tolerance for diversity. Furthermore, these outcomes were mediated by the quality of the placement, the level of reflection in the classroom, and the overall relevance of the course content to the students' placement. Driscoll et al. (1996) reported similar findings in their preliminary study of 4 courses (spanning four disciplinary areas), although they measured students' sensitivity to diversity.

Of 3 studies that included both surveys and open-ended questions, the study by Osborne et al. (1998) stands out because of its methodological rigor. In particular, this study included a randomized experimental and control group in which subjects were not informed about the inclusion of a service component until after they had already enrolled in the course. Four sections of a health care course (two experimental and two control), which included 95 subjects (57 females), were investigated to determine the impact of service on students' ability to work with diverse others and their awareness of diversity issues. Although there were no pretest differences among groups, there were dramatic changes in posttest scores between the experimental and control groups. Quantitative results revealed significant increases in the experimental groups' ability to work with diverse others, whereas the control group showed decreases along this measure. In addition, qualitative analyses showed marked increases in the experimental groups' awareness of diversity compared to the control group. Although this study minimized threats to both internal and external validity, questions remain as to what aspects of the course were most beneficial (e.g., reflection, group discussion) and whether there were differential effects across racial groups.

The studies by Rice and Brown (1998) and Giles and Eyler (1994) reinforce earlier findings that service raises awareness, reduces stereotypes, increases knowledge and understanding, and enhances communication and interpersonal skills in relation to difference. Rice and Brown, however, relied on student evaluations, which are susceptible to demand characteristics, to substantiate their qualitative conclusions. Giles and Eyler's results also warrant caution because more than 80% of their sample was involved in previous service activities, no control group was used, and qualitative change was assessed more from posttest descriptions than from pre-post change. Their results, however, suggest that service may have cumulative benefits for students who engage in multiple service activities over time. There are certainly limitations to all of the mixed-method studies reviewed, but these studies seem to point overwhelmingly to the effectiveness of service interventions to improve students' racial attitudes, awareness, understanding, and knowledge.

Overall Effectiveness of Service Interventions

In general, the weight of the evidence points to the effectiveness of service interventions in reducing racial bias. However, only 1 study (Dunlap, 1998) examined effects across racial groups, finding markedly fewer effects on students of color than on White students. Although service-learning was found to be a more effective modality in 1 study (Myers-Lipton, 1996), the majority of studies combined different service programs into one group, making it difficult to reach any definitive conclusions about the most effective approach. In addition, 2 studies (Myers-Lipton, 1996; Astin et al., 1999) demonstrated the long-term sustainability of service interventions; and 2 studies (Eyler et al., 1997a; Astin & Sax, 1998) found similar effects across different disciplinary areas. Several studies also pointed to both reflection and interaction as important underlying processes to reduce bias, while other studies suggested that placement quality mediates the effectiveness of service interventions. Overall, there seems to be a general consistency across studies indicating that service programs are effective in reducing aspects of students' racial bias. But many gaps are evident in the extant knowledge base that prevent generalizing these results to all student groups and institutions.

Overall Effectiveness

Table 2 presents the overall effects across the higher education curriculum and co-curriculum. Of the 73 reported findings from different intervention studies, 52 were positive, 14 were mixed, and 7 were nonsignificant. Thus, taken at face value, the weight of the evidence suggests that the majority of educational interventions are effective in reducing racial bias. As a whole, differential effects were found most often across different racial groups. In particular, White students were found to benefit more than students of color, especially across diversity workshop and training interventions. Similarly, gender effects were found in a small number of studies, with women usually experiencing stronger effects or making greater change during the course. Class-level effects were found among a handful of studies, although the multicultural intervention studies reported conflicting effects and several noted the tendency for seniors to report higher level of biases than freshman. Long-term effects were consistent across several studies, with a range of 6 months to 5 years post-intervention. Finally, and perhaps the least explored, are the disciplinary effects. Only two studies found significant differences across different academic disciplines,

TABLE 2
Summary of overall effects for educational interventions

	General education				
Type of effect	Multicultural course intervention	Diversity workshop or training intervention	Peer-facilitated intervention	Service intervention	Row total
Positive	17	9	14	12	52
Mixed	7	1	5	1	14
No change	5	1	1	0	7

with the social and behavioral sciences reporting higher gains than students in business or the natural sciences. Unfortunately, the magnitude of these effects is unclear, as the majority of studies did not test for differential effects.

To make a substantive evaluation of the effectiveness of these interventions, both the strengths and limitations of the studies must be weighed against the findings. With regard to overall strengths, these studies represent a broad range of methodological approaches and research designs. The balance between quantitative, qualitative, and mixed-method approaches allows for a more detailed portrait of the processes and characteristics that underlie student change. Several studies also incorporated sophisticated research designs that control for both internal and external threats to validity. And the range of national and single-institution studies, coupled with the varied survey and quasi-experimental research designs, offer an illuminating perspective on both the predictive and the absolute nature of change effects. The national studies, in particular, allow for some generalizability to other colleges and universities. In addition, several of the qualitative studies provide important insight into the theoretical underpinnings of various interventions. Finally, the inclusion of mixed-method studies allows for an in-depth exploration of how the various interventions influence different aspects of students' racial bias.

Despite the strengths of many of these studies, in the majority of cases, their limitations cast doubt on the evidentiary weight of the findings. In general, the limitations fall into four major categories: *conceptualization, measurement, research design,* and *analytic approach.* With regard to conceptualization, many studies neglected to ground their studies in a theoretical framework that considered prior research on racial bias. Furthermore, few studies offered formalized hypotheses that made explicit the links between their proposed intervention and the intended student outcomes. As a result, only a small number of studies, mainly qualitative in design, investigated which aspects of an intervention were most influential in enacting change. In addition, few studies hypothesized or investigated what underlying processes were involved in translating the various interventions into reductions in racial bias.

With regard to measurement, many studies relied on instruments that lack sufficient conceptual clarity and rigor to measure racial bias. The national studies, in particular, often base their findings on secondary data analysis, using measures that were not necessarily intended to assess racial bias. This shortcoming is further complicated by the inherent measurement error in using single-item measures to tap into complex phenomena. In addition, few studies included social desirability scales or measures of racial bias designed to tap into more contemporary forms of racial bias (e.g., Modern Racism Scale). As such, many of the measures are prone to ceiling effects or demand characteristics that distort the true intervention effects.

The limitations most prevalent across the various studies were related to the research design. First, the vast majority of studies that relied on quasi-experimental designs did not control for nonequivalency across experimental and control groups, with the result that they were vulnerable to selection bias. Second, the majority of the studies relied on either convenience or purposive sampling, which severely limited generalizability to other populations. This limitation was exacerbated in many qualitative studies that focused exclusively on students who demonstrated marked improvements. Third, given the overall short duration of many of the studies, few included design elements (e.g., posttest-only group) to measure the

presence of a testing bias. Fourth, many of the national and quasi-experimental studies neglected to control for confounding variables that could suppress or veil the true intervention effects, especially among studies without pretest or control groups. Fifth, few studies incorporated plans to test the sustainability of effects over time. Sixth, in most cases, samples were aggregated across race, gender, and academic disciplines, a practice that often distorts the true effects of an intervention, especially in studies that report overall positive findings.

With regard to analytic approaches, many of the quantitative studies relied on simple *t* tests to assess overall course effects. This technique does not permit the researcher to control for confounding variables, and unless equivalency checks are included in the design, the results are often misleading and inflated; a similar critique applies to studies using ANOVA techniques. Qualitative studies are also limited by the use of single raters to develop coding schemes and analyze data. In addition, many of the qualitative studies relied on single sources of data, often student assignments or course evaluations, which are tarnished by instructor demand characteristics. Finally, few of the qualitative or mixed-method studies triangulated data or relied on member checking to further substantiate their claims.

Given the numerous limitations found throughout this review, the overall quality of the majority of the studies is called into question. Moreover, although the weight of the evidence leans in the positive direction, 22 of the 74 findings suggest that many of the educational interventions produce mixed or nonsignificant results. And the small number of studies that investigated differential effects reported inconsistent findings across various groups. Thus no definitive statements can be made about the true effectiveness of the various interventions reviewed.

Nevertheless, as a collection these studies provide important clues about needed improvements and gaps in the extant knowledge base. In the remaining sections, recommendations are offered to improve the quality of the intervention studies, to better evaluate change in racial bias, and to identify questions that remain for future investigation.

IMPROVING THE QUALITY OF INTERVENTION STUDIES

Based on the limitations raised in the previous section, this section presents suggestions for improving the overall quality of the research and evidence across four main areas: conceptualization, measurement, research design, and analytic approach.

Conceptualization

More attention must be paid to drawing theoretical links between various interventions and racial bias. In most of the reviewed studies, researchers suggest that there is a linear path between a particular intervention and racial bias outcomes. But a linear picture of the nature of change is far too simplistic. Rather, there are many underlying psychological processes that mediate change, and these need to be taken into consideration when conceptualizing any study investigating racial bias.

Researchers in social psychology, for instance, have theorized a number of cognitive, affective, and behavioral processes that are responsible for translating educational approaches into changes in racial bias. Gaertner and Dovidio (2000), for instance, posit that content-based knowledge elicits change in racial bias through the process of decategorization, whereby students break down group

barriers and conceive of themselves and others as separate individuals. Gaertner and Dovidio also suggest that sociohistorical knowledge of various groups often decreases uncertainty and discomfort (Gaertner & Dovidio, 1986) while changing individuals' perceptions of what is appropriate and just (Duckitt, 1992). In addition, research suggests that interventions enact change by increasing empathy (Stephan & Finlay, 1999), reducing negative feelings (Dovidio & Gaertner, 1998; Stephan & Stephan, 1985) and arousing self-directed negative emotion (Devine & Monteith, 1993; Dovidio & Gaertner, 1998). A long line of research on intergroup contact (see Allport, 1954; Pettigrew, 1998b; Pettigrew & Tropp, 2000) also suggests a number of factors that must be present for contact between groups to reduce bias: equal status between groups, cooperative as opposed to competitive intergroup interaction, opportunities for personal acquaintance, and supportive norms by authorities both within and outside the contact situation. Thus increased attention to the conceptual links that bridge various interventions and intergroup bias will provide a more theoretically grounded approach to understanding student change.

Attention to the psychological processes that mediate racial bias is crucial, but so, too, is the role of developmental effects in fostering change. A number of reviewed studies (cf. Stake & Hoffman, 2001; Bidell et al., 1994) implicate both cognitive and identity development as important processes that influence racial bias outcomes. In addition, several studies have demonstrated more general links between cognitive complexity and racial bias (Guthrie, 1996; Taylor, 1994; Gurin et al., 2002), and between racial identity development and racial bias (Taylor, 1990; Tatum, 1992; Helms, 1990). In particular, these studies suggest that changes in racial bias are dependent on students' level of cognitive and identity development. For instance, the early stages of racial identity are often associated with less tolerant attitudes toward outgroup members. Thus models that are conceptually grounded in developmental theory may capture the indirect effects of various interventions and help to explain why a particular intervention was ineffective.

Measurement

In addition, more attention needs to be directed at the instruments used to measure racial bias. First, studies rely overwhelmingly on attitudinal measures of prejudice to assess changes in racial bias. Additional attention, however, is needed in measuring other aspects of racial bias, such as stereotypes, affective reactions, and discrimination. By incorporating multiple dimensions of racial bias in a particular study, researchers will gain a better understanding of the overall impact of their interventions and lessen the likelihood that ceiling effects will distort their results. Second, more studies need to include social desirability scales for a better understanding of the reliability of self-report measures. Instruments such as the Modern Racism Scale (McConahay, 1986), which includes social desirability scales, are a good start, although modified versions are needed that assess attitudes other than those directed toward African Americans. Finally, in keeping with the conceptual recommendations, measures are needed that tap into the underlying processes that mediate student change. A number of instruments that measure intergroup anxiety (Stephan & Stephan, 1985), empathy (Davis, 1983), and contact (Gurin et al., 1999) are available and would add greatly to our knowledge of how various interventions influence racial bias.

Research Design

Many research design enhancements are needed to improve the overall quality of intervention studies:

1. More effort must be spent on minimizing problems due to nonequivalency. Randomization in group selection is optimal but is rarely an option. Instead, researchers need to include pretest measures and take necessary steps to ensure that there are no differences between experimental and control groups. When differences are found, either a matched control group or interaction measures along significant variables are needed to control for selection effects. Instructor characteristics are also important to consider, as research has shown differences in instructional practices across racial groups (Hurtado, 2001).
2. More efforts are needed to control for possible testing effects, especially in short-term studies. The Solomon four-group design is the ideal choice for quantitative studies, but using measures with proven test-retest reliability is a strong alternative.
3. Designs need to include long-term assessments of student change. Such designs are certainly more labor intensive and resource intensive, but they carry considerable weight in providing evidence of the sustainability of course effects.
4. Qualitative designs should address key design elements, including the specific type of design (e.g., grounded theory design, phenomenological design, or case study design); reflections on the researcher's role; data collection, recording, and analysis procedures; verification steps; and a clear delineation of narrative outcomes (see Creswell, 1994, for an elaboration of these areas).
5. More efforts need to be directed at minimizing demand characteristics. Although administering instruments outside of class is ideal, alternative steps, such as using individuals other than the instructor to administer testing, would aid considerably in reducing these confounding effects.
6. Better coordination among quantitative and qualitative measures used in mixed-method studies would enhance the translation fidelity and ensure an adequate system of checks and balances.

Analytic Approach

Finally, more rigorous analytic approaches are needed to better explain the true intervention effects:

1. Studies need to disaggregate students of color into separate racial group categories. Small sample sizes certainly limit statistical power, but researchers whose studies include national samples should be more mindful that minority groups are often differentially affected by educational interventions.
2. With regard to quantitative studies, more attention is needed in examining structural paths among independent and dependent variables. Both path analysis and structural equation modeling are viable alternatives; the latter allows for the simultaneous estimation of hypothesized regressions while controlling for measurement error.

505

3. Qualitative researchers need to pay more attention both to both rival expla-
nations and to the data for students who showed little or no improvement.
Doing so can help them determine why some students are more affected than
others and how other variables outside the classroom may be influencing stu-
dent change.

4. Data analysis in qualitative studies needs to rely on both multiple raters
(especially individuals less involved with the study) and multiple forms of
data. Multiple sources of data allow the researcher to rule out rival explana-
tions while adding needed validity to the findings in many studies.

EVALUATING EDUCATIONAL INTERVENTIONS

Figure 1 provides a temporal schematic of the change process based on the find-
ings, strengths, limitations, and improvements discussed in previous sections of
the review. The framework is not meant to imply causal associations, as many of the
paths remain untested. Rather, it is meant to highlight those areas in need of con-
sideration when conducting an evaluation of educational interventions. Depending
on the particular methodology employed in a study (i.e., quantitative, qualitative,
or mixed-method), certain areas will take on greater significance.

Beginning with the far-left block in the diagram, both student and institutional
characteristics are highlighted as important pre-intervention controls. These mea-
sures take on particular relevance in quantitative studies that survey a broad range
of institutions. For instance, numerous studies (e.g., Gurin et al., 2002; Taylor,
1994; Milem, 1994) suggest that both structural diversity and the overall climate
for diversity at an institution can influence students' racial bias. Furthermore,
although conflicting results were found regarding the cumulative nature of various
interventions, prior exposure to diversity courses or diverse peers has been shown
to influence racial bias outcomes in several of the national studies (cf. Antony,
1993; Hyun, 1994; Milem, 1994). In addition, background variables related to race,
gender, and socioeconomic factors are important variables to consider when con-
trolling for potential confounding factors. For example, several studies (e.g., Chang,
2002) demonstrated interaction effects based on background characteristics. The
next part of the diagram includes pretest controls for both racial bias and students'
cognitive and identity development. As noted in the "Overall Effectiveness" section,
without a baseline measure of students' entering level of racial bias, it is impossi-
ble to disentangle confounding effects related to selection and history biases.
Moreover, such an omission limits the researchers' ability to understand the nature
of change derived from the intervention itself. In addition, developmental pretests
are included on the basis of the improvements noted in the previous section. Under-
standing students' baseline level of cognitive and identity development can help
researchers to detect the indirect effects of an intervention, while adding concep-
tual clarity to the nature of student change.

Moving to the center block in the diagram, important aspects of the interven-
tion itself are highlighted to remind researchers to probe deeper into the many
nuances of an intervention that may be responsible for enabling change. Few stud-
ies, for instance, examined the environmental effects on a particular intervention,
leaving many questions unanswered about how instructor characteristics, racial
makeup of the class, class level, or discipline influenced the nature of change. The
remaining features of the intervention pertain to the typology presented earlier in

Pre-Intervention Controls

Pretest Measures

Educational Intervention

Intermediary Processes

Posttest Measures

Student characteristics
- Background
- Diversity exposure

Institutional characteristics
- Structural diversity
- Climate for diversity

Racial bias (Time 1)
- Prejudice
- Stereotypes
- Affective reactions
- Discrimination

Development (Time 1)
- Cognitive
- Identity

Environment
- Instructor characteristics
- Racial makeup
- Class level
- Discipline

Goals
- Implicit versus explicit

Pedagogical focus
- Content versus contact

Educational processes
- Didactic instruction
- Experiential learning

Duration
- Short versus long

Psychological processes
- Cognitive
- Affective
- Behavioral

Racial bias (Time 2)
- Prejudice
- Stereotypes
- Affective reactions
- Discrimination

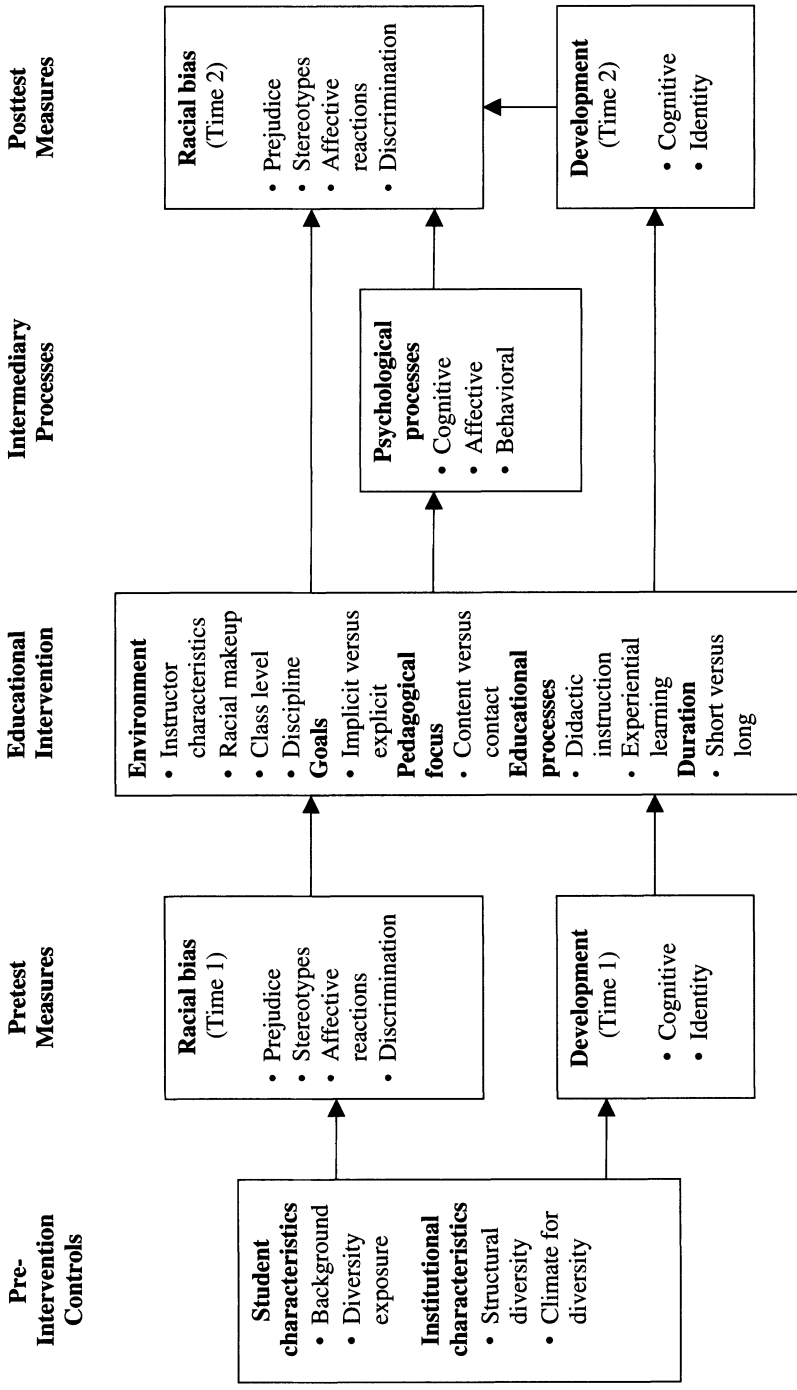**Development (Time 2)**
- Cognitive
- Identity

FIGURE 1. *Conceptual framework for evaluating the influence of educational interventions on racial bias.*

507

the review. Researchers need to be mindful of how the goals, pedagogical focus, educational processes, and duration of an intervention influence the change process. Many of these factors remain unexplored, and few studies have incorporated a comparative design to determine the benefits of different approaches.

The next block refers to the intermediary processes discussed in the previous section. Future intervention studies need to examine more thoroughly how various interventions work to reduce racial bias. By including measures that tap into the underlying psychological processes that link various interventions to racial bias, researchers will gain a better understanding of which processes translate which types of interventions into changes in racial bias. In addition, these processes may shed light on why a particular intervention failed to cause change.

The far-right section of the diagram includes posttest measures of both racial bias and cognitive and identity development. By using repeated outcome measures, researchers will have the ability to isolate change that occurred during the intervention. Although not included in the diagram, post-posttest measures can also be used to better determine the long-term sustainability of the intervention effects.

THE FUTURE OF INTERVENTION STUDIES

This final section of the review explores gaps in the extant knowledge base. In particular, it is intended to raise questions that remain unanswered, based on the numerous studies reviewed in this article. These questions fall into four main areas: *differential effects, comparative effects, long-term effects,* and *elemental effects.*

Differential Effects

A number of questions remain regarding who benefits from various types of educational interventions. Although several studies examined effects across racial groups, no definitive conclusions were reached regarding the efficacy of interventions to benefit all groups. More research is needed to explain why some interventions are more efficacious for certain racial groups and why other groups may be more resistant to participating in particular types of interventions. Moreover, few studies examined how gender, class level, and disciplinary differences affected the success rates of various interventions. In particular, little evidence was presented about how interactions among these variables influenced the nature of change. If reducing racial bias among all student groups remains an important goal of higher education, more attention is needed to determine how exposure to various educational interventions is related to different normative peer contexts.

Additional questions also remain about the influence of institutional factors on the efficacy of various interventions. Although the national studies included a representative group of institutions, few compared institutional types to explain differences in the predictive nature of interventions. Furthermore, although many of the national studies controlled for both structural diversity and the climate for diversity, few compared programs across those dimensions. The overall climate for diversity, as well as a particular university's commitment to promoting diversity, may influence students' willingness to engage more seriously in the issues raised in an intervention. Answers to these questions will shed light on whether

institutional factors play a predominant role in influencing the efficacy of different interventions.

Few studies examined the comparative effects of various types of interventions. For instance, questions remain as to whether content-based interventions are more effective in reducing prejudice or whether contact-based interventions are better suited to change behavioral patterns and social distance. In general, little is known about the magnitude of the effects across types of interventions or whether certain interventions are more effective at targeting certain types of bias. To answer these questions, more studies are needed that include a comprehensive plan to investigate multiple types of interventions. By using similar measures across multiple interventions, researchers will gain a better understanding of the comparative effects across programs. In addition, comparative designs may illuminate whether a particular pedagogical focus (i.e., content versus. contact) produces consistent effects regardless of the larger programmatic design.

A few studies looked at the long-term effects of various interventions, but many questions remain about the overall sustainability of program effects, as well as the cumulative nature of those effects. For instance, it may be that particular interventions are better equipped to make lasting impressions on students and therefore lasting changes. The duration of a program may also influence the sustainability of its effects. And repeating a similar intervention may produce different cumulative effects than does a particular sequence of interventions that are not similar to each other. Answers to these questions will help administrators and practitioners to decide whether a one-course diversity requirement is sufficient to meet the university's goals for educating a diverse student body.

Finally, many questions remain unanswered with regard to how the various elements of a particular intervention actually work to enact change. The racial makeup or size of the class and the instructor or facilitator's race, gender, or level of experience may influence student outcomes and the nature of change. Further research is needed to explore whether particular types of assignments, readings, discussions, or group projects impart more change than others. Although classes that create structured opportunities for interaction seem important, change may depend more on the quality of interactions than their frequency. The sequencing of course elements and other educational processes may also be important factors to consider, especially by researchers interested in understanding why some interventions are more effective with certain types of students. Answers to these questions may aid instructors and facilitators in designing a particular intervention and may provide more clues about why some interventions are more effective than others.

In conclusion, many questions remain about the links between educational interventions and racial bias. In most cases, research is at an incipient stage of development, with much room for improvement and expansion. Although this review arrived at no definitive conclusions regarding the overall efficacy of various interventions, the positive nature of many of the findings suggests great potential in programs to reduce racial bias. Given the tenuous nature of intergroup relations on many campuses today, there is a strong need to continue to investigate the efficacy of various educational interventions. Hopefully, this review will serve as a guide to researchers on how to strengthen and improve their studies while illuminating the path of future exploration.

References

Adams, P. (1973). *Attitude exploration survey.* Amherst, MA: University of Massachusetts Center for Racial Understanding.

Adorno, T. W., Frenkel-Brunswick, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality.* New York: Harper & Row.

Alimo, C., Kelly, R., & Clark, C. (2002). Diversity initiatives in higher education: Intergroup dialogue programs, student outcomes, and implications for campus racial climate. *Multicultural Education, 10*(1), 49–53.

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

Antony, J. (1993, November). *Can we all get along? How college impacts students' sense of the importance of promoting racial understanding.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Pittsburgh. (ERIC Document Reproduction Service No. ED 365174).

Astin, A. W. (1993a). Diversity and multiculturalism on the campus. How are students affected? *Change, 23,* 44–49.

Astin, A. W. (1993b). *What matters in college? Four critical years revisited.* San Francisco: Jossey-Bass.

Astin, A. W., Astin, H. S., Bayer, A. E., & Bisconti, A. S. (1997). Overview of the unrest period. In L. F. Goodchild & H. S. Wechsler (Eds.), *The history of higher education* (pp. 724–738). Boston: Simon & Schuster Custom Publishing.

Astin, A. W., & Sax, L. J. (1998). How undergraduates are affected by service participation. *Journal of College Student Development, 39*(3), 251–263.

Astin, A. W., Sax, L. J., & Avalos, J. (1999). Long-term effects of volunteerism during the undergraduate years [Electronic version]. *Review of Higher Education, 22*(2), 187–202.

Banks, J. A. (2001). *Cultural diversity and education: Foundations, curriculum, and teaching* (4th ed.). Boston: Allyn & Bacon.

Bell, L. A., & Griffin, P. (1997). Designing social justice education courses. In M. Adams, L. A. Bell, & P. Griffin (Eds.), *Teaching for diversity and social justice: A source book.* New York: Routledge.

Bennett, C. I. (1999). *Comprehensive multicultural education: Theory and practice* (4th ed.). Boston: Allyn & Bacon.

Bidell, T. R., Lee, E. M., Bouchie, N., Ward, C., & Brass, D. (1994, November). *Developing conceptions of racism among young White adults in the context of cultural diversity coursework.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 377270)

Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 259–374). Mahwah, NJ: Erlbaum.

Blimling, G. S. (1993). The influence of college residence halls on students. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (9th ed., pp. 248–307). New York: Agathon.

Bobo, L., & Kluegel, J. R. (1997). Status, ideology, and dimensions of Whites' racial beliefs and attitudes: Progress and stagnation. In P. A. Katz & D. A. Taylor (Eds.), *Eliminating racism: Profiles in controversy* (pp. 85–114). New York: Plenum Press.

Bobo, L., Kluegel, J. R., & Smith, R. A. (1997). Laissez-faire racism: The crystallization of a kinder, gentler, antiblack ideology. In S. A. Tuch & J. K. Martin (Eds.), *Racial attitudes in the 1990s: Continuity and change* (pp. 15–42). Westport, CT: Praeger.

Bodenhausen, G. V., Sheppard, L. A., & Kramer, G. P. (1994). Negative affect and social judgment: The differential impact of anger and sadness. *European Journal of Social Psychology, 24,* 45–62.

Brehm, C. J. (1998, September). *Stereotypes, tolerance, and the classroom.* Paper presented at the annual meeting of the American Political Science Association, Boston. (ERIC Document Reproduction Service No. ED 426927)

Brower, A. M., & Ketterhagen, A. (2004). Is there an inherent mismatch between how Black and White students expect to succeed in college and what their colleges expect from them? *Journal of Social Issues, 60*(1), 95–116.

Brown, C. R., & Mazza, G. J. (1991). Peer training strategies for welcoming diversity. In J. C. Dalton (Ed.), *Racism on campus: Confronting racial bias through peer interventions* (pp. 39–52). San Francisco: Jossey-Bass.

Brown, R. (1965). *Social psychology.* New York: Free Press.

Brown, R. (1995). *Prejudice: Its social psychology.* Malden, MA: Blackwell.

Bruffee, K. A. (1999). *Collaborative learning: Higher education, interdependence, and the authority of knowledge* (2nd ed.). Baltimore, MD: John Hopkins University Press.

Byrnes, D. A., & Kiger, G. (1988). Contemporary measures of attitudes toward Blacks. *Educational and Psychological Measurement, 48,* 107–118.

Cabrera, A. F., & Nora, A. (1994). College student perceptions of prejudice and discrimination and their feelings of alienation: A construct validation approach. *Review of Education/Pedagogy/Cultural Studies, 16*(3–4), 387–409.

Cabrera, A. F., Nora, A., Bernal, E. M., Terenzini, P. T., & Pascarella, E. T. (1998, November). *Collaborative learning: Preferences, gains in cognitive and affective outcomes, and openness to diversity among college students.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Miami. (ERIC Document Reproduction Service No. ED 427589)

Cabrera, A. F., Nora, A., Crissman, J. L., Terenzini, P. T., Bernal, E. M., & Pascarella, E. T. (2002). Collaborative learning: Its impact on college students' development and diversity. *Journal of College Student Development, 43*(1), 20–34.

Chang, M. J. (2002). The impact of an undergraduate diversity course requirement on students' racial views and attitudes. *Journal of General Education, 51*(1), 21–42.

Cockrell, K. S., Caplow, J. A., & Donaldson, J. F. (2000). A context for learning: Collaborative groups in the problem-based learning environment. *Review of Higher Education, 23,* 347–364.

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches.* Thousand Oaks, CA: Sage Publications.

Dalton, J. C. (1991a). Racial and ethnic backlash in college peer culture. In J. C. Dalton (Ed.), *Racism on campus: Confronting racial bias through peer interventions* (pp. 3–12). San Francisco: Jossey-Bass.

Dalton, J. C. (Ed.). (1991b). *Racism on campus: Confronting racial bias through peer interventions.* San Francisco: Jossey-Bass.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44,* 113–126.

Devine, P. G. (1989). Stereotypes and prejudice: The automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18.

Devine, P. G., & Monteith, M. J. (1993). The role of discrepancy-associated affect in prejudice reduction. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in intergroup perception* (pp. 317–344). Orlando, FL: Academic Press.

Diaz, R., Furco, A., & Yamada, H. (1996). *Student pre-post service assessment.* Berkeley, CA: Service-Learning Research & Development Center, Graduate School of Education, University of California.

Dollard, J., Doob, L., Miller, N. E., Mowrer, O., & Sears, R. (1939). *Frustration and aggression.* New Haven, CT: Yale University Press.

Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice, and discrimination: Another look. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 276–319). New York: Guilford.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about survey questions: Meaning, memory, attitudes, and social interaction* (pp. 204–237). New York: Russell Sage Foundation.

Dovidio, J. F., & Gaertner, S. L. (1998). On the nature of contemporary prejudice: The causes, consequences, and challenges of aversive racism. In J. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 3–32). Newbury Park, CA: Sage.

Dovidio, J. F., Gaertner, S. L., Stewart, T. L., Esses, V. M., Vergert, M., & Hodson, G. (2004). From intervention to outcome: Processes in the reduction of bias. In W. G. Stephan & W. P. Vogt (Eds.), *Education programs for improving intergroup relations* (pp. 243–265). New York: Teachers College Press.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2003). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 175–197). Malden, MA: Blackwell.

Driscoll, A., Holland, B., Gelmon, S., & Kerrigan, S. (1996). An assessment model for service-learning: Comprehensive case studies of impact on faculty, students, community, and institution. *Michigan Journal of Community Service Learning* (Fall), 66–71.

Duckitt, J. (1992). Psychology and prejudice. *American Psychologist, 47*(10), 1182–1193.

Dunlap, M. R. (1998). Voices of students in multicultural service-learning settings. *Michigan Journal of Community Service Learning* (Fall), 58–67.

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 269–322). New York: McGraw-Hill.

Esses, V. M., Haddock, G., & Zanna, M. P. (1993). Values, stereotypes, and emotions as determinants of intergroup attitudes. In D. M. Mackie & D. L. Hamilton (Eds.), *Affect, cognition, and stereotyping: Interactive processes in group perception* (pp. 137–166). New York: Academic Press.

Etzioni, A. (1983). *An immodest agenda.* New York: McGraw Hill.

Everett, K. D. (1998). Understanding social inequality through service learning. *Teaching Sociology, 26,* 299–309.

Eyler, J., & Giles, D. E. (1999). *Where's the learning in service learning?* San Francisco: Jossey-Bass.

Eyler, J., Giles, D. E., & Braxton, J. (1997a). The impact of service-learning on college students. *Michigan Journal of Community Service Learning* (Fall), 5–15.

Eyler, J., Giles, D. E., & Braxton, J. (1997b, March). *Report of a national study comparing the impacts of service-learning program characteristics on post secondary students.* Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 408505)

Fishbein, M., & Azjen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.

Forgas, J. P., & Moylan, S. J. (1991). Affective influences on stereotype judgments. *Cognition and Emotion, 5,* 379–395.

Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Orlando, FL: Academic Press.

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The Common Ingroup Identity Model.* Philadelphia, PA: Psychology Press.

Gaff, J. G. (1991). *New life for the college curriculum.* San Francisco: Jossey-Bass.

Geranios, C. A. (1997). Cognitive, affective, and behavioral outcomes of multicultural courses and intergroup dialogues in higher education. *Dissertation Abstracts International, 58* (03), 764A. (UMI No. 9725284)

Giles, D. E., & Eyler, J. (1994). The impact of college community service laboratory on students' personal, social, and cognitive outcomes. *Journal of Adolescence, 17,* 327–339.

Gurin, G. (1992). *The Michigan Study: Expectations and experiences of first-year students with diversity.* Ann Arbor, MI: University of Michigan, Office of Academic and Multicultural Initiatives.

Gurin, P., Dey, E. L., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review, 72*(3), 330–367.

Gurin, P., & Marcus, H. (1988). Group identity: The psychological mechanisms of durable salience. *Revue Internationale de Psycologie Sociale, 1*(2), 257–274.

Gurin, P., Miller, A. H., & Gurin, G. (1980). Stratum identification and consciousness. *Social Psychology Quarterly, 43,* 30–47.

Gurin, P., Nagda, B. A., & Lopez, G. E. (2004). The benefits of diversity in education for democratic citizenship. *Journal of Social Issues, 60*(1), 17–34.

Gurin, P., Peng, T., Lopez, G., & Nagda, B. A. (1999). Context, identity, and intergroup relations. In D. Prentice & D. Miller (Eds.), *Cultural divides: Understanding and overcoming group conflict* (pp. 133–170). New York: Russell Sage Foundation.

Guthrie, V. L. (1996). The relationships of levels of intellectual development and levels of tolerance for diversity among college students. *Dissertation Abstracts International, 57* (11), 4666A. (UMI No. 9713104)

Gutiérrez, L. M. (1989). Ethnic consciousness, consciousness raising and the empowerment process of Latinos. *Dissertation Abstracts International, 51* (01), AAF9013915.

Hamilton, D. L., & Rose, T. (1980). Illusory correlation and the maintenance of stereotype beliefs. *Journal of Personality and Social Psychology, 39,* 832–845.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review, 103,* 336–255.

Hamilton, D. L., & Troiler, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127–163). Orlando, FL: Academic Press.

Hasslen, R. C. (1993). The effects of teaching strategies in multicultural education on monocultural college students' perceptions. *Dissertation Abstracts International, 55* (01), 56A. (UMI No. 9413010)

Hathaway, R. S., III. (1999). College student definitions of key concepts and thinking in a multicultural education course. *Dissertation Abstracts International, 60* (07), 2401A. (UMI No. 9938385)

Haugsby, T. (1991). Service learning and cooperative education: Serving God or Mammon. *Journal of Cooperative Education, 27*(2), 99–106.

Helms, J. (1990). *Black and White racial identity: Theory, research, and practice.* Westport, CT: Praeger.

Henderson-King, D., & Kaleta, A. (2000). Learning about social diversity: The undergraduate experience and intergroup tolerance. *Journal of Higher Education, 71*(2), 142–164.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53,* 575–604.

Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology, 47,* 237–271.

hooks, b. (1994). *Teaching to transgress: Education as the practice of freedom.* New York: Routledge.

Hughes, M. (1994). Helping students understand and appreciate diversity. In C. Schroeder & P. Mable (Eds.), *Realizing the educational potential of residence halls* (pp. 190–217). San Francisco: Jossey-Bass.

Humphreys, D. (1997). *General education and American commitments: A national report on diversity course and requirements.* Washington, DC: Association of American Colleges and Universities.

Humphreys, D. (2000). National survey finds diversity requirements common around the country [Electronic version]. *Diversity Digest, 5*(1).

Hurtado, S. (2001). Linking diversity and educational purpose: How diversity affects the classroom environment and student development. In G. Orfield (Ed.), *Diversity challenged: Evidence on the impact of affirmative action.* Cambridge, MA: Harvard Education Publishing Group.

Hurtado, S., Carter, D. F., & Spuler, A. (1996). Latino student transition to college. *Research in Higher Education, 37*(2), 135–157.

Hyun, M. (1994, November). *Helping to promote racial understanding: Does it matter if you're Black or White?* Paper presented at the annual meeting of the Association for the Study of Higher Education, Tucson. (ERIC Document Reproduction Service No. ED 375710)

Inkelas, K. K. (1998, November). *Increasing visibility: The racial attitudes of Asian Pacific American students.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Miami.

Johnson, D. W., & Johnson, R. T. (2000). The three C's of reducing prejudice and discrimination. In S. Oskamp (Ed.), *Reducing discrimination and prejudice* (pp. 239–268). Mahwah, NJ: Lawrence Erlbaum.

Jones, J. M. (1972). *Prejudice and racism* (1st ed.). Reading, MA: Addison-Wesley.

Katz, D., & Braly, K. (1933). Racial stereotypes in one hundred college students. *Journal of Abnormal and Social Psychology, 28,* 280–290.

Katz, J. H., & Ivey, A. (1977). White awareness: The frontier of racism awareness training. *Personnel and Guidance Journal* (April), 485–489.

Khan, S. R. (1999). Teaching an undergraduate course on the psychology of racism. *Teaching of Psychology, 26*(1), 28–33.

Kluegel, J. R. (1990). Trends in Whites' explanations of the Black–White gap in SES. *American Sociological Review, 55,* 512–525.

Lawrence, S. M. (1998). Unveiling positions of privilege: A hands-on approach to understanding racism. *Teaching of Psychology, 25*(3), 198–200.

LeVine, R. A., & Campbell, D. T. (1972). *Ethnocentrism: Theories of conflict, ethnic attitudes, and group behavior.* New York: John Wiley.

Lopez, G. E. (1993). The effects of group contact and curriculum on White, Asian American and African American students' attitudes. *Dissertation Abstracts International, 54* (07), 3900B. (UMI No. 9332125)

Lopez, G. E., Abboushi, M., & Reifman, A. (1992). *Interest in social and political thinking.* Unpublished manuscript, University of Michigan, Ann Arbor.

Lopez, G. E., Gurin, P., & Nagda, B. A. (1998). Education and understanding structural causes for group inequalities. *Political Psychology, 19*(2), 305–329.

Maas, A., Castelli, L., & Arcuri, L. (2000). Measuring prejudice: Implicit versus explicit techniques. In D. Capozza & R. J. Brown (Eds.), *Social identity processes: Trends in theory and research* (pp. 79–121). Beverly Hills, CA: Sage Publications.

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology, 79*(4), 602–616.

Mackie, D. M., & Smith, E. R. (1998). Intergroup relations: Insights from a theoretically integrative approach. *Psychological Review, 105*(3), 499–529.

MacPhee, D., Kreutzer, J. C., & Fritz, J. J. (1994). Infusing a diversity perspective into human development courses. *Child Development, 65,* 699–715.

Markus, G. B., Howard, J., & King, D. (1993). Integrating community service and classroom instruction enhances learning: Results from an experiment. *Educational Evaluation and Policy Analysis, 15,* 410–419.

Marin, P. (2000). The educational possibility of multi-racial/multi-ethnic college classrooms. In American Council on Education & American Association of University Professors, *Does diversity make a difference? Three research studies on diversity in college classrooms.* Washington, DC: Authors.

McCauley, C., Wright, M., & Harris, M. E. (2000). Diversity workshops on campus: A survey of current practices at U.S. colleges and universities [Electronic version]. *College Student Journal, 34*(1), 1–12.

McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic Press.

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? *Journal of Conflict Resolution, 25*(4), 563–579.

Meyers-Lipton, S. J. (1996). Effect of a comprehensive service-learning program on college students' level of modern racism. *Michigan Journal of Community Service Learning* (Fall), 44–54.

Milem, J. F. (1991, November). *The role of college peer groups and faculty reference groups in the development of students' attitudes toward race.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Boston. (ERIC Document Reproduction Service No. ED 339321)

Milem, J. F. (1994). College, students, and racial understanding. *Thought and Action, 9*(2), 51–92.

Miller, W. E., Kinder, D. R., & Rosenstone, S. J. (1993). *American National Election Study, 1992: Pre- and post-election survey.* Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Miller, W. E., & Miller, A. H. (1977). *The CPS 1976 National Election Study.* Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Mullen, B., & Hu, L. T. (1989). Perception of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology, 10*(3), 233–252.

Nagda, B. A., Gurin, P., & Lopez, G. (2003). Transformative pedagogy for democracy and social justice. *Race Ethnicity and Education, 6,* 165–191.

Nagda, B. A., Kim, C., & Truelove, Y. (2004). Learning about difference, learning with others, learning to transgress. *Journal of Social Issues, 60*(1), 195–214

Nagda, B. A., Spearmon, M. L., & Holley, L. C. (1999). Intergroup dialogues: An innovative approach to teaching about diversity and justice in social work programs [Electronic version]. *Journal of Social Work Education, 35*(3), 433–449.

Nagda, B. A., & Zúñiga, X. (2003). Fostering meaningful racial engagement through intergroup dialogues. *Group Processes & Intergroup Relations, 6*(1), 115–132.

Nelson, L., Johnson, C., Boyd, J., & Scott, T. (1994, April). *Effects of participation in an intergroup communication program: An assessment of Shippenburg University's Building Bridges program.* Paper presented at the annual meeting of the Eastern Psychological Association, Providence, RI. (ERIC Document Reproduction Service No. ED 370152)

Neville, H., & Furlong, M. (1994). The impact of participation in a cultural awareness program on the racial attitudes of first-year college students. *Journal of College Student Development, 35,* 371–377.

Nora, A., & Cabrera, A. F. (1996). The role of perceptions of prejudice and discrimination on the adjustment of minority students to college. *Journal of Higher Education, 67*(2), 119–148.

Oakes, P. J., Haslam, S. A., & Turner, J. C. (1994). *Stereotyping and social reality.* Cambridge, MA: Blackwell.

Osborne, R. E., Hammerich, S., & Hensley, C. (1998). Student effects of service-learning: Tracking change across a semester. *Michigan Journal of Community Service Learning* (Fall), 5–12.

Palmer, B. (2000). The impact of diversity courses: Research from Pennsylvania State University. *Diversity Digest* (Winter). Retrieved January 17, 2002, from *http://www.diversityweb.org/digest/W00/research.html*

Pascarella, E. T., Edison, M., Nora, A., Hagedorn, L. S., & Terenzini, P. T. (1996). Influence on students' openness to diversity and challenge in the first year of college. *Journal of Higher Education, 67*(2), 174–195.

Pascarella, E. T., Ethington, C. A., & Smart, J. C. (1988). The influence of college on humanitarian/civic involvement values. *Journal of Higher Education, 59,* 412–437.

Pascarella, E. T., & Terenzini, P. T. (1980). Student-faculty and student-peer relationships as mediators of the structural effects of undergraduate residence arrangement. *Journal of Educational Research, 73,* 344–353.

Pascarella, E. T., & Terenzini, P. T. (1981). Residence arrangement, student/faculty relationships, and freshman-year educational outcomes. *Journal of College Student Personnel, 22,* 147–156.

Pascarella, E. T., Terenzini, P. T., & Blimling, G. S. (1994). The impact of residential life on students. In C. Schroeder & P. Mable (Eds.), *Realizing the educational potential of residence halls* (pp. 22–52). San Francisco: Jossey-Bass.

Pence, D. J., & Fields, J. A. (1999). Teaching about race and ethnicity: Trying to uncover White privilege for a White audience. *Teaching Sociology, 27,* 150–158.

Pettigrew, T. F. (1998a). Prejudice and discrimination on the college campus. In J. L. Eberhardt & S. T. Fiske (Eds.), *Confronting racism: The problem and the response* (pp. 263–279). Thousand Oaks, CA: Sage.

Pettigrew, T. F. (1998b). Intergroup contact theory. *Annual Review of Psychology, 49*(1), 65–85.

Pettigrew, T. F., & Tropp, L. R. (2000). Does intergroup contact reduce prejudice? Recent meta-analytic findings. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 93–114). Mahwah, NJ: Erlbaum.

Pike, G. R. (2002). The differential effects of on- and off-campus living arrangements on students' openness to diversity. *NASPA Journal, 39*(4), 283–299.

Ramirez, M., III (1991). *Psychotherapy and counseling with minorities: A cognitive approach to individual and cultural differences.* New York: Pergamon Press.

Rhoads, R. A. (1998). In the service of citizenship: A study of student involvement in community service. *Journal of Higher Education, 69*(3), 183–203.

Rice, K. L., & Brown, J. R. (1998). Transforming educational curriculum and service learning. *Journal of Experiential Education, 21*(3), 140–146.

Rokeach, M., Smith, P., & Evans, R. (1960). Two kinds of prejudice or one? In M. Rokeach (Ed.), *The open and the closed mind* (pp. 132–168). New York: Basic Books.

Romanoff, S. J. (2000). The learning community laboratory: A context for discovery. *Journal of College Student Development, 41*(2), 245–247.

Ryan, C. S., Park, B., & Judd, C. M. (1996). Assessing stereotype accuracy: Impli-cations for understanding the stereotyping process. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 121–157). New York: Guilford.

Saltmarsh, J., & Heffernan, K. (2000). *Introduction to service-learning: Readings and resources for faculty.* Providence, RI: Campus Compact Publications.

Schoem, D., Hurtado, S., Sevig, T., Chesler, M., & Sumida, S. H. (2001). Intergroup dialogue: Democracy at work in theory and practice. In D. Schoem & S. Hurtado (Eds.), *Intergroup dialogue: Deliberative democracy in school, college, community, and workplace.* Ann Arbor, MI: University of Michigan Press.

Schuman, H., Steech, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations.* Cambridge, MA: Harvard University Press.

Sears, D. O. (1988). Symbolic racism. In P. A. Katz & D. A. Taylor (Eds.), *Eliminat-ing racism: Profiles in controversy* (pp. 53–84). New York: Plenum Press.

Shapiro, N. S., & Levine, J. H. (1999). *Creating learning communities.* San Francisco: Jossey-Bass.

Sherif, M., Harvey, O. J., White, B. J., Hood, W. R., & Sherif, C. W. (1961). *Inter-group conflict and cooperation: The Robbers Cave experiment.* Norman: University of Oklahoma Press.

Sidanius, J. (1993). The psychology of group conflict and the dynamics of oppression: A social dominance perspective. In S. Iyengar & W. J. McGuire (Eds.), *Explorations in political psychology* (pp. 183–219). Durham, NC: Duke University Press.

Sidanius, J., Devereux, E., & Pratto, F. (1992). A comparison of symbolic racism the-ory and social dominance theory as explanations for racial policy attitudes. *Journal of Social Psychology, 132*(3), 377–395.

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hier-archy and oppression.* Cambridge, UK: Cambridge University Press.

Slavin, R. E. (1995). Cooperative learning groups and intergroup relations. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural edu-cation* (pp. 628–634). New York: Macmillan.

Sleeter, C. E., & Grant, C. A. (1999). *Making choices for multicultural education: Five approaches to race, class, and gender.* Upper Saddle River, NJ: Merrill.

Smith, E. R., & Ho, C. (2002). Prejudice as intergroup emotion: Integrating relative deprivation and social comparison explanations of prejudice. In I. Walker & H. J. Smith (Eds.), *Relative deprivation: Specification, development, and integration* (pp. 332–348). New York: Cambridge University Press.

Smith, K. M. (1993, May). The impact of college on White students' racial attitudes. Paper presented at the annual meeting of the Association for Institutional Research, Chicago. (ERIC Document Reproduction Service No. ED 360923)

Solorzano, D., Ceja, M., & Yosso, T. (2000). Critical race theory, racial microaggres-sions, and campus racial climate: The experiences of African American college stu-dents. *Journal of Negro Education, 69*(1–2), 60–73.

Spencer, M. S., & Nagda, B. A. (2002, August). *Outcomes of intergroup dialogues for students in a graduate-elective and undergraduate required course.* Paper presented at the annual meeting of the American Psychological Association, Chicago.

Springer, L., Palmer, B., Terenzini, P. T., Pascarella, E. T., & Nora, A. (1996). Atti-tudes toward campus diversity: Participation in a racial or cultural awareness work-shop. *Review of Higher Education, 20*(1), 53–68.

Stake, J. E., & Hoffman, F. L. (2001). Changes in student social attitudes, activism, and personal confidence in higher education: The role of women's studies [Electronic version]. *American Educational Research Journal, 38*(2), 411–436.

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111,* 42–61.

Steckler, G. (1957). Authoritarian ideology in Negro college students. *Journal of Abnormal Social Psychology, 54,* 396–399.

Stephan, W. G. (1985). Intergroup relations. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3, pp. 599–658). New York: Addison-Wesley.

Stephan, W. G., & Finlay, K. (1999). The role of empathy in improving intergroup relations. *Journal of Social Issues, 55,* 729–743.

Stephan, W. G., & Stephan, C. W. (1985). Intergroup anxiety. *Journal of Social Issues, 41*(3), 157–175.

Stephan, W. G., & Stephan, C. W. (2001). *Improving intergroup relations.* Thousand Oaks, CA: Sage.

Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago: Nelson.

Tatum, B. D. (1992). Talking about race, learning about racism: The application of racial identity development theory in the classroom. *Harvard Educational Review, 62*(1), 1–24.

Taylor, K. E. (1990). The dilemma of difference: The relationship of the intellectual development, racial identity, and self-esteem of Black and White students to their tolerance for diversity. *Dissertation Abstracts International, 52* (02), 443A. (UMI No. 9121439)

Taylor, S. H. (1994). Enhancing tolerance: The confluence of moral development with the college experience. *Dissertation Abstracts International, 56* (01), 114A. (UMI No. 9513291)

Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *Journal of Higher Education, 68*(6), 599–623.

Vanman, E. J., Paul, B. Y., Ito, T. A., & Miller, N. (1997). The modern face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology, 73,* 941–959.

Vasques Scalera, C. M. (1999). Democracy, diversity, dialogue: Education for critical multicultural citizenship. *Dissertation Abstracts International, 61* (02), 561A. (UMI No. 9959879)

Vogelgesang, L. J. (2001, November). *The impact of college on the development of civic values: How do race and gender matter?* Paper presented at the annual meeting of the American Educational Research Association, Seattle. (ERIC Document Reproduction Service No. ED 451791)

Vogt, P. V. (1997). *Tolerance and education: Learning to live with diversity and difference.* London: Sage.

Vohra, S., Rodolfia, E., De La Cruz, A., Vincent, C., & Bee-Gates, D. (1991). A cross-cultural training format for peer counselors. *Journal of College Student Development, 32,* 82–84.

Walt Whitman Center for the Culture and Politics of Democracy. (1993). *The Civic Skills Test: A critical examination working group notebook.* New Brunswick, NJ: Author.

Whitt, E. J., Edison, M. I., Pascarella, E. T., Terenzini, P. T., & Nora, A. (2001). Influences on students' openness to diversity and challenge in the second and third years of college. *Journal of Higher Education, 72*(2), 172–204.

Wilder, D., & Simon, A. F. (2003). Affect as a cause of intergroup bias. In R. Brown & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes* (pp. 153–172). Malden, MA: Blackwell.

Wolford, K., & Clemo, L. (1997, March). *Collaborative and experiential learning in a human diversity course.* Paper presented at the Annual Conference on Undergraduate Teaching of Psychology, Ellenville. (ERIC Document Reproduction Service No. ED 448642)

Yeakley, A. M. (1998). The nature of prejudice change: Positive and negative change processes arising from intergroup contact experiences. *Dissertation Abstracts International, 59* (10), 3970A. (UMI No. 9910030)

Zanna, M. P., & Rempel, J. K. (1988). Attitudes: A new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 315–334). Cambridge, UK: Cambridge University Press.

Zúñiga, X., & Nagda, B. A. (1993). Dialogue groups: An innovative approach to multicultural learning. In D. Schoem, L. Frankel, X. Zúñiga, & E. Lewis (Eds.), *Multicultural teaching in the university* (pp. 233–248). Westport, CT: Praeger.

Zúñiga, X., & Nagda, B. A., & Sevig, T. D. (2002). Intergroup dialogues: An educational model for cultivating engagement across differences. *Equity & Excellence in Education, 35*(1), 7–17.

Zúñiga, X., Nagda, B. A, Sevig, T. D., Thompson, M., & Dey, E. (1995, November). *Speaking the unspeakable: Student learning outcomes in intergroup dialogues on a college campus.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Orlando, FL.

Author

MARK E. ENGBERG was a Research Associate at the University of Michigan, School of Education, while writing this article. He can be reached at 4545 N. Seeley Ave., Apt. 1, Chicago, IL 60625; e-mail *mengberg@umich.edu.* His research focuses on the educational benefits of diversity in postsecondary institutions and the role of educational interventions in improving intergroup relations.

APPENDIX
Measures used to assess racial bias

Study	Type	Research design	Measures
Brehm (1998)	Multicultural	Quantitative	Stereotype measure (questions on 7-point scale) in which subjects were asked to rate groups on various characteristics (wealth, work ethic, intelligence, dependency, violence); tolerance levels toward civil liberties of different groups. (Both measures were created by the author.)
Chang (2002)	Multicultural	Quantitative	An 8-item adaptation of the Modern Racism Scale (McConahay et al., 1981).
Henderson-King & Kaleta (2000)	Multicultural	Quantitative	A feelings thermometer used in studies that evaluate different social groups by examining intergroup and intragroup attitudes and prejudice (Miller & Miller, 1974); Beliefs About Racism (created by the authors).
Hurtado (2001)	Multicultural	Quantitative	Acceptance of People of Different Races/Cultures; Cultural Awareness; Tolerance of People with Different Beliefs (1987–1991 Cooperative Institutional Research Program student survey).
Inkeles (1998)	Multicultural	Quantitative	Attitudes Toward Affirmative Action Policies and Practices (items derived from 1990–1994 Michigan Student Survey; Gurin, 1992).
Lopez (1993)	Multicultural	Quantitative	Awareness of Racial Inequality (items derived from 1990 Michigan Student Survey; Gurin, 1992).
Palmer (2000)	Multicultural	Quantitative	Attitudinal measures (not described).
Smith (1993)	Multicultural	Quantitative	Attitudes Toward Affirmative Action Practices (6-item scale derived from a 1990 research study at the University of Michigan).
Taylor (1994)	Multicultural	Quantitative	Tolerance Scale (openness to human difference—attitudes toward race, gender, and sexual orientation; items derived from 1990–1994 Michigan Student Survey; Gurin, 1992).
Hathaway (1999)	Multicultural	Qualitative	Classroom observations, student interviews, student papers, and weekly questions.
Khan (1999)	Multicultural	Qualitative	Course evaluations and written assignments.
Marin (2000)	Multicultural	Qualitative	In-depth interview with faculty, focus groups with students and faculty, classroom observation, document reviews of course syllabi, and student evaluations.
Hasslen (1993)	Multicultural	Mixed	Cultural Diversity Awareness Inventory; Situational Attitude Scale (Whites' attitudes toward Blacks); journals and student writings (see Hasslen, 1993, for additional details).

Bidell et al. (1994)	Multicultural	Mixed	Written questionnaire (2 questions on nature of racism and causes of racism).
MacPhee et al. (1994)	Multicultural	Mixed	Old-Fashioned Racism Scale and Modern Racism Scale (McConahay, 1986).
Stake & Hoffman (2001)	Multicultural	Mixed	General Egalitarian Attitudes Toward Others, adapted from Moral Values Scale (Miller et al., 1993); General Awareness of Discrimination Against Others, based on Equal Opportunities/Rights Scale (Miller et al., 1993); subjective change measure (authors' creation).
Astin (1993a)	Multicultural; diversity workshop and training	Quantitative	Cultural Awareness; Commitment to Promoting Racial Understanding (CIRP); Discrimination Is No Longer a Problem (see Astin, 1993b, for an elaboration of measures).
Gurin et al. (2002)	Multicultural; diversity workshop and training	Quantitative	Racial/Cultural Engagement (index of items representing self-change in awareness, appreciation, and acceptance of persons from different races, derived from 1985–1989 CIRP student survey; single-item "learned about other race/ethnic groups" derived from 1990–1994 Michigan Student Survey; Gurin, 1992).
Milem (1994); Antony (1993); Vogelgesang (2001); Hyun (1994)	Multicultural; diversity workshop and training	Quantitative	CIRP 1985 Freshman Survey; CIRP 1989 Follow-up Survey; CIRP 1993–1996 Student Information Form; 1998 CIRP College Student Survey.
Katz & Ivey (1977)	Diversity workshop and training	Mixed	Steckler Anti-Black and Anti-White Inventory (covers a variety of negative stereotyped opinions toward both Blacks and Whites; Steckler, 1957); Attitude Exploration Survey (used to assess White people's awareness of racism; Adams, 1973); behavioral measures, students' journals, and follow-up questionnaires.
Neville & Furlong (1994)	Diversity workshop and training	Quantitative	Social Scale and Social Scenarios Scale (Byrnes & Kiger, 1988).
Pascarella et al. (1996); Whitt et al. (2001)	Diversity workshop and training	Quantitative	Openness Toward Diversity/Challenge (National Study of Student Learning [NSSL]).
Springer et al. (1996)	Diversity workshop and training	Quantitative	Openness Toward Diversity (NSSL).

(continued)

521

APPENDIX (*Continued*)

Study	Type	Research design	Measures
Nelson et al. (1994)	Peer-facilitated	Quantitative	Ten statements (e.g., "Affirmative action is reverse discrimination"; "Most minorities are not qualified to be here"; "Minorities contribute to learning") evaluated on a 7-point Likert scale (1 = *disagree*, 7 = *strongly agree*).
Pence & Fields (1999)	Peer-facilitated	Qualitative	Instructors' impressions.
Pascarella & Terenzini (1980, 1981)	Living-learning	Quantitative	Personal development (e.g., openness to new ideas, developing interpersonal skills, and better understanding of self).
Pike (2002)	Living-learning	Quantitative	Openness Toward Diversity (NSSL).
Geranios (1997)	Intergroup dialogue	Quantitative	Survey instrument designed by ASU Voices of Discovery Program, which included measures of students' knowledge of and stereotypes/attitudes toward diverse groups.
Gurin et al. (2004)	Intergroup dialogue	Quantitative	Non-Divisiveness of Difference; Perception of Commonalities in Values Across Groups; Mutuality in Learning About Own and Other Groups (items derived from the Michigan Student Survey and the Intergroup Relations, Conflict, and Community Program at the University of Michigan; see Gurin, 1992, for information about the Michigan Student Survey).
Gurin et al. (1999)	Intergroup dialogue	Quantitative	Group Identity; Intergroup Perceptions (intergroup divisiveness, commonality of interests, positive/negative views of conflict); Intergroup Emotions (positive/negative intergroup emotion); Approval/Disapproval of University Diversity Policies (i.e., attitudes toward affirmative action and multicultural policies); majority of items derived from two senior-year questionnaires administered by the Michigan Student Survey (Gurin, 1992) and the University of Michigan Psychology Department.
Lopez et al. (1998)	Intergroup dialogue	Quantitative	Causal Attributions of Racial or Ethnic Inequality (Gurin, Miller, & Gurin, 1980); Causal Analysis of Intergroup Conflict (students asked to write about causes of conflict; Gutiérrez, 1989).

Study	Approach	Method	Measures
Nagda & Zúñiga (2003)	Intergroup dialogue	Quantitative	Critical Social Awareness (e.g., Importance of Race [Gurin & Markus, 1988], Centrality of Race [Gurin & Markus, 1988]; Dialogic Thinking (e.g., Comfort in Communicating Across Difference [Zúñiga, Nagda, Sevig, Thompson, & Dey, 1995]); Building Bridges (Zúñiga et al., 1995).
Nagda et al. (2003)	Intergroup dialogue	Quantitative	Socio-Historical Thinking; Structural Thinking About Race Inequality; Social Structural Understanding (items derived from the Michigan Student Survey and the Intergroup Relations, Conflict and Community Program at the University of Michigan; see Gurin, 1992, for information about the Michigan Student Survey).
Nagda et al. (2004)	Intergroup dialogue	Quantitative	Importance of and Confidence in Reducing Prejudice and Promoting Diversity (authors' creation).
Spencer & Nagda (2002)	Intergroup dialogue	Quantitative	Comfort in Communicating Across Difference (Zúñiga et al., 1995); Cognitive Centrality of Race (Gurin & Marcus, 1988; Importance of Race (Gurin & Marcus, 1988; Critical Social Awareness (Fletcher, 1986; Lopez, Abboushi, & Reifman, 1992); Building Bridges (Zúñiga et al., 1995).
Alimo et al. (2002)	Intergroup dialogue	Qualitative	Interviews.
Vasques Scalera (1999)	Intergroup dialogue	Qualitative	Analysis of questionnaires, in-depth interviews, and student reflection papers.
Yeakley (1998)	Intergroup dialogue	Qualitative	Semi-structured interviews.
Nagda et al. (1999)	Intergroup dialogue	Mixed	Focus groups, surveys, in-depth interviews with 10 students, and observations.
Cabrera et al. (2002); Cabrera et al. (1998)	Collaborative learning	Quantitative	Openness Toward Diversity (NSSL).
Wolford & Clemo (1997)	Collaborative learning	Quantitative	Multicultural Experience Inventory (Ramirez, 1991).

APPENDIX (*Continued*)

Study	Type	Research design	Measures
Lawrence (1998)	Collaborative learning	Qualitative	Reflective writings and class observations.
Astin et al. (1999); Astin & Sax (1998)	Service-learning	Quantitative	CIRP 1985, 1990–1994 Freshman Survey; CIRP 1989, 1994–1995 Follow-up Survey
Everett (1998)	Service-learning	Quantitative	Course evaluation form
Eyler et al. (1997a, 1997b)	Service-learning	Quantitative	Tolerance items (respecting views of others and empathetic to all points of views; developed as part of the "Measuring Citizens Project" [Whitman Center, 1993])
Myers-Lipton (1996)	Service-learning	Quantitative	Modern Racism Scale (McConahay, 1986)
Dunlap (1998)	Service-learning	Qualitative	Reflective journals
Rhoads (1998)	Service-learning	Qualitative	Formal and informal interviews, surveys, participatory observation, and document analysis
Driscoll et al. (1996)	Service-learning	Mixed	In-person assessments (interviews, focus groups, and document analysis); independent reflection measures (journals, pre-post surveys); review of documentation (syllabi, activity reports)
Eyler & Giles (1999)	Service-learning	Mixed	Multi-item survey (tolerance for diversity, negative stereotypes, and appreciation for other cultures)
Giles & Eyler (1994)	Service-learning	Mixed	Questionnaire at week 1, 5, 13 and open-ended questions related to awareness and stereotyping (items derived from multiple sources including Markus, Howard, & King 1993; Pascarella, Ethington, & Smart, 1988; and Astin, 1993b)
Osborne et al. (1998)	Service-learning	Mixed	Self-Perception Scale (measures students' perceived ability to work with diverse others; created by first author); written work coded for awareness of diversity
Rice & Brown (1998)	Service-learning	Mixed	Student Service Assessment (Diaz, Furco, & Yamada, 1996); course evaluation