The Impact of a Program of Prejudice-Reduction Seminars in South Africa

M. Alan McCool, Jr.
Queen’s University Belfast
Belfast, Northern Ireland

Fanie Du Toit
Institute for Justice and Reconciliation
Cape Town, Republic of South Africa
and
Stellenbosch University
Stellenbosch, Republic of South Africa

Christopher R. Petty
Stellenbosch University
Stellenbosch, Republic of South Africa

Clark McCauley
Solomon Asch Center for Study of Ethnopolitical Conflict
Philadelphia, PA

The influence of weekend-long prejudice-reduction seminars on attitudes, knowledge, and behavioral expectations of South African student leaders was assessed quantitatively and qualitatively. Pre–post comparisons indicated increases in perceived stereotyping-caused suffering, ease of committing cross-cultural mistakes, and comfort with approaching racially diverse strangers. The seminars did not reduce essentialist thinking, nor the in-group favoritism found with measures of group identification and comfort approaching strangers. Similarly, they did not affect summary measures of stereotype understanding, perceived inter- or intra-group heterogeneity, preferred model of interethnic relations, nor attitudes toward affirmative action. Despite limited evidence of seminar impact, participants evaluated the seminars highly. Qualitative results suggest that opportunities to address racially based misperception of self and in-group were particularly valued.

1This research was supported by a postdoctoral fellowship from the Solomon Asch Center for Study of Ethnopolitical Conflict and by a grant from the United States Institute of Peace. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the United States Institute of Peace or the Solomon Asch Center for Study of Ethnopolitical Conflict. The authors extend their thanks to Ifat Maoz for her very helpful comments on an earlier draft of this report.

2Correspondence concerning this manuscript should be addressed to Clark McCauley, Solomon Asch Center for Study of Ethnopolitical Conflict, University of Pennsylvania, St. Leonard’s Court Suite 305, 3819-33 Chestnut Street, Philadelphia, PA 19104. E-mail: cmccaul@psych.upenn.edu
Since the end of the Cold War, symbolized by the demolition of the Berlin Wall in 1989, intergroup violence has been associated largely with ethnic or communal conflicts (Schlesinger, 1997). In some places (e.g., Israel/Palestine), the violence continues apace. In other places (e.g., South Africa, East Timor, Northern Ireland), a degree of stability has been achieved, and the challenge is to consolidate the security of the general population and to build the peace.

In many ethnically diverse countries, including the United States, a popular approach to peace building and enhancing organizational performance is to bring together individuals from different segments of society for an educational program intended to encourage more accurate intergroup perceptions, more positive intergroup feelings, and more cooperative intergroup behavior. These intergroup contact interventions are variously identified as workshops, programs, seminars, or training, and these nomina
tives usually are qualified by prefixes such as sensitivity, consciousness-raising, diversity (e.g., McCauley, Wright, & Harris, 2000), cultural-awareness (Neville & Furlong, 1994), or prejudice- or stereotype-reduction (Louw-Potgieter, Kamfer, & Boy, 1991). The interventions are implemented in a great variety of forms, but generally rest on the theoretical base of Allport’s (1954) contact hypothesis, which has been a source of inspiration and controversy in social psychology for decades (Forbes, 1997; Pettigrew, 1998; Pettigrew & Tropp, 2000; Stephan, 1987).

As contact interventions come in a great variety, so does the empirical work that relates amount of contact with an out-group to attitude toward that out-group. For instance, a great many survey studies have correlated reported past contact with current attitude (e.g., Hewstone et al., 2005; Towles-Schwen & Fazio, 2001; Wood & Soleitner, 1996; for a review, see Pettigrew & Tropp, 2000). There are also laboratory studies of exercises in intergroup tolerance that involve majority participants only (e.g., Byrnes & Kiger, 1990), and a considerable number of studies that focus on programs designed for children (for a review, see Bigler, 1999).

Stephan and Stephan (2001) recently examined the range of contact interventions and focused on those, usually called diversity initiatives or diversity training, that are used most commonly in educational and corporate settings. Diversity initiatives generally are aimed at highlighting the differences in experience and ways of life between various ethnic groups. Newly cognizant of the differences between groups, participants are then encouraged to appreciate more fully these differences and the out-group members who display them. Diversity initiatives, particularly those that are focused on valuing as opposed to managing diversity, are commonly quite short, lasting only hours or days. They generally include both didactic and
experiential exercises and may emphasize the feelings produced by experiencing stereotyping, bias, or discrimination.

As concern over preparing individuals for multiethnic social environments has grown, the popularity of diversity initiatives has burgeoned (Rynes & Rosen, 1995; Stephan & Stephan, 2001). Evaluation of the impact of these interventions, however, has lagged far behind their practice (Neville & Furlong, 1994; Noe & Ford, 1992; Stephan & Stephan, 2001). Participants often are queried at or near the close of a seminar regarding their interest in its content, their satisfaction with its administration and facilitators, and their overall opinion of the value of the experience.

While helpful in planning and improving seminar administration, this process evaluation cannot substitute for product evaluation: evidence that the intervention produced a change in the participants (e.g., in their attitudes toward out-groups) that advances the goals of intergroup harmony and reconciliation. Product evaluations of diversity initiatives have been rare. In their review, Stephan and Stephan (2001) were able to locate only seven. Of the examined programs, only three were judged to have been largely successful. One showed no lasting effect at all, and the remaining three indicated unfavorable results, as well as favorable ones.

It should be noted that in just the past few years, a number of new and innovative evaluation studies have been conducted, often engaging adult participants and showing some success in altering attitudes (Gurin, Peng, Lopez, & Nagda, 1999; Maoz, 2000, 2002; Nagda & Zuniga, 2003; for a review, see Nevo & Brem, 2002). These studies have concentrated on dialogue groups, a form of intervention that continues over days or even months of relatively unstructured, small-group conversations between participants from ethnic groups in violent conflict. These dialogic techniques show great promise for informing future practice, and the discourse-based approach to their evaluation (Maoz, 2004; also see Steinberg & Bar-On, 2002) offers the potential for important insight into key reconciliation processes. Compared with the diversity initiatives common in education and corporate settings, however, dialogue groups are longer in duration, less structured with prepared exercises, and more focused on current intergroup violence.

The program evaluated in the present research is a diversity initiative that purposefully brings adult members of ethnic groups with a history of conflict into face-to-face contact for a variety of structured intergroup interactions addressing the relationship between the groups under the leadership of facilitators trained for the purpose. In addition to being of the sort common in business settings, the program is distinguished by an exceptional social context. It was carried out in South Africa, where the requirement for effective intergroup conciliation interventions is acute. South Africa is also
of interest because its interethnic conflicts involve not just two groups, but many. One residue of the apartheid government in South Africa is the existence of at least the following commonly recognized groups: Whites (usually either of British heritage or Afrikaans-speaking), Indians, Coloureds (persons of mixed or Southeast Asian heritage\(^3\)), and African Blacks (divided, to some extent, by precolonial identities). Participants from these backgrounds were recruited from the Student Representative Councils and related governing bodies of several large South African universities.

Quantitative measures of participants’ beliefs, attitudes, and understanding were made both before and after the seminar under the general hypothesis that program completion would reduce disparities in beliefs regarding different race groups, improve attitudes toward out-groups, and provide a better understanding of the process and function of stereotyping. A number of qualitative measures were included as well. These served to assess behavioral intentions after the workshop, to provide a certain degree of process evaluation, and to identify potential effects of the program that were unanticipated and, therefore, were unrepresented among the quantitative items.

Method

Participants

Participants were recruited from the student leadership of three major South African universities, which will be identified as U1, U2, and U3. Workshop organizers sought students who were influential on campus so that the effects of the intervention might have the widest possible impact. Thus, the Student Representative Council (SRC) of each university was approached with the offer of “free leadership and diversity training.” Each council had approximately 15 members, so the invitation also was extended to about 20 additional prominent students at each university. In all instances, the offer was greeted with enthusiasm, and participants were

\(^3\)As do other authors (e.g., Jones, 1997), we recognize that, although racial categories may be unfounded and hurtful, the groups commonly identified as racial groups can nonetheless enjoy considerable social recognition. It thus becomes very difficult to discuss the related issues without using the commonly applied labels. Given this necessity, we note that the term Coloured is not generally used in South Africa as a name for the whole community of people with mixed or Southeast Asian heritage. However, the associated adjective (i.e., coloured) is used commonly to describe individuals from these groups. Because there is no widely accepted, succinct alternative, and because it is grammatically interchangeable with Black and White, we capitalize Coloured in this work, retain its British spelling, and use it as a proper noun for a grouping with quite a broad membership.
recruited easily. Each seminar was held at an off-campus conference facility, at which all participants were required to spend the weekend.

A separate seminar was conducted for each university’s students between March and August 2001. There were 26 students from U1. When they were asked to provide their “former ‘racial’ classification,” 19 specified Black, 3 White, 1 Coloured, 2 Indian, and 1 gave his race as Malay. (For the sake of brevity, former “‘racial’ classification” will be referred to simply as race.) There were 37 participants from U2: 13 Black, 16 White, 6 Coloured, and 2 who did not give their race. Finally, there were 21 participants from U3: 9 Black, 6 White, 5 Coloured, and 1 Indian. Of the 84 participants, 45 were men and 39 were women. All of the participants were 18 to 30 years old (Mdn = 21.4 years for the three groups, inclusive).

Research Design

The simplest approach to measuring change is a pretest–posttest design in which the same participants are assessed before and after the intervention of interest; in this case, the prejudice-reduction seminar. Taken alone, this design can produce misleading results if the pretest sensitizes or prepares participants in a way that leads them to react differently to the intervention than participants who have not been pretested (Cook & Campbell, 1979). In order to assess this possibility, a split design was used in which only half of the participants were pretested, but all participants received the posttest.

Slightly more than half of the registered participants in each seminar were selected at random to complete a pretest version of the evaluation questionnaire. Because a few of those who registered for the seminar would inevitably be unable to participate, pretesting slightly more than half of registrants assured that at least half of participants would be pretested. Pretests were distributed to designees in the days immediately before the seminar. They were accompanied by instructions to complete the questionnaire individually and to turn it in upon arrival at the seminar.

Of the 47 participants who eventually completed the pretest, 72% indicated in the posttest that they had never before taken part in any similar seminar, 8% were unsure, and 19% indicated that they had. Similar percentages were obtained for the posttest-only participants: 65%, 5%, and 30%, respectively; $\chi^2(2, N = 84) = 1.42$, ns. Previous participation was related to two of the substantive measures, as described in the Results.

Seminar

The seminar was presented in three phases: an introductory, more theoretical phase; a dialogical, more affective phase; and a skills-focused,
The conative phase. Each seminar was facilitated by two or three individuals who varied in both gender and race. The facilitator team did not have the same members at each of the three seminars, but continuity was provided by the consistent presence of one experienced individual. The team for each seminar also included a recent graduate, similar in age to the participants and often with SRC experience. Facilitators were staff members of the Institute for Justice and Reconciliation of Cape Town, who had received on-the-job training in facilitation skills.

The first phase of the program lasted for the 3 hours of the initial evening session, typically a Friday evening. It included introductions, the formulation of values to act as guides during discussion, and a short theoretical description of stereotyping. The analogy of a mask, which perceivers figuratively place on the faces of group members, was used. Imposition of mask-like stereotypical perception was discussed as preventing the appreciation of the characteristics of the individuals being perceived.

The description of stereotyping was followed by a playful introduction to how stereotypes function in society. The interactive exercises and exploratory discussions of this period focused on less emotive issues, such as stereotypes of fashion, membership in organizations, and generational differences.

The projected outcomes of the seminar’s first phase were predominantly process-oriented. These included achieving some measure of personal commitment and group cohesion among participants, establishing a secure forum for candid and constructive dialogue, and alerting participants to the importance of stereotyping as a social phenomenon. One product outcome of Phase 1 was to communicate some basic theory on the role of stereotyping in intergroup relations.

The second, affective phase of the seminar lasted for most of the second day (a Saturday) and consisted of four sessions. The four sessions demonstrated the relationships between stereotyping on the one hand and dialogue, history, racial identity, and personal narratives, respectively, on the other hand. More specifically, the first session comprised a set of experiential exercises that show how labeling can impede dialogue and teamwork.

In the second session, participants discussed each others’ personal histories and perspectives, superimposed on a timeline of nationally important South African democratization events. In the third session, a key event on the seminar agenda, participants were encouraged to produce and discuss a list of traits associated with each of the racial groupings substantially represented in the seminar. In the fourth and final session, participants were given the chance, in a small-group format, to relate how they had each been affected by and had dealt with the South African transition from apartheid to democracy.

Overall, the second phase of the seminar was designed to start on a lighter note, and gradually progress to a serious engagement with the effect
of racial stereotyping on reconciliation and justice in South Africa. Projected outcomes of this phase include greater awareness of how stereotyping can impede dialogue and of how historical perspectives differ among South African youth, depending on their background. At the same time, this phase illustrated human suffering in every racial group, drawing attention to the influence of the South African transition on personal identity construction, and consequently to the implications of negative stereotyping between racial groupings.

The third, skills-focused phase of the seminar occurred on the third day (Sunday) and was comprised of three parts. The first part was a consideration of the previous day, focusing on three questions. The first question is “Why do people use stereotypes?” Participants considered the stereotypical labels produced the previous day, and were encouraged to try to identify the origins of particular associations. Participants also were encouraged to identify stereotypes based on partial or complete misunderstanding or ignorance, and those that are particularly hurtful. The second question is “What (in situations of direct interracial contact and given the reality of negative intergroup stereotyping) does one person expect of another?” The final question is “Given that negative stereotyping can be hurtful, misleading, and may impede dialogue, why do people persist with stereotyped views of others?”

The second part of the seminar on Sunday involved a deliberate attempt at a transfer of practical skills that could help individuals to challenge stereotyping in their own thinking and environment. It was structured, again, by three leading questions: namely, “How do I challenge negative stereotyping in my own thinking?”; “How do I challenge negative stereotyping directed at myself?”; and “How do I challenge stereotyping in my context?”

The last part of the seminar was an attempt to bring closure to the weekend through the compilation of a contact list for seminar participants and the preparation, by a committee chosen by the participants, of a declaration of commitment. The declaration was signed by each willing participant with a brief comment on the gains of the weekend. The seminar concluded with the ritual burning of the lists of racially stereotypical traits and the completion of the posttest questionnaire.

*Questionnaire*4

Here, we provide a brief overview of its structure. Part 1 of the questionnaire (“About you”) contained demographic items. Part 2 (“Fill-In

---

4A copy of the questionnaire is available from the authors.
Items”) asked for traits associated with different races. The fill-in items were not included in the posttest for U2 or U3, and the numbers of subsequent parts of the questionnaire were correspondingly shifted down by one. The change was made because at the end of the U1 seminar, it became clear that listing stereotypical traits in the posttest conflicted too starkly with the symbolism of burning the trait lists immediately beforehand.

Part 3 of the questionnaire (“True or False”) was comprised of 10 items devised to measure socially desirable responding. The items failed to cohere into a usable scale, however, and will not be discussed further.

Part 4 was labeled “Opinion Items.” Opinion items were answered on a 7-point scale ranging from −3 (disagree strongly) to 3 (agree strongly) to indicate extent of agreement with the accompanying statement. In order to facilitate comparison with other items, scores on opinion items are represented in the Results section on a scale ranging from 1 to 7.

Part 5 (“Judgment Scales”) was the largest part of the questionnaire. Several open-ended items and two fill-in-the-blank items appeared in Part 5, but responses to most items were made on a 7-point scale ranging from 1 to 7. Only the endpoints of each scale were labeled, and the labels implied a unipolar metric. The endpoints labeled as 1 identified anchors such as none, not great, or no change; while the endpoints labeled as 7 identified anchors such as much, very great, or great change.

Part 6 (“Final Items”) contained items, in varying response formats, that addressed summative questions regarding the seminar program as a whole. In the posttest, this final part included retrospective items inappropriate to, and therefore not appearing in, the pretest.

In addition to questionnaire responses, the notes of the first author, who acted as a participant–observer during the first workshop (U1), were available for consideration. Questionnaire responses of the author and two other participant–observers in the U1 seminar were not included in the data set.5

Results

Our findings are reported in three main sections. The first describes responses to items in which participants explicitly provided their own, subjective evaluations of seminar processes and outcomes. The second examines participants’ acquisition of new information that seminar organizers specifically aimed to communicate. The third looks at how some of the participants’ beliefs about ethnic differences were affected by the seminar experience.

5Findings arising from participant observations will appear where most relevant to questionnaire responses.
Analysis

Items on the evaluation questionnaires generally were included as constituents of different a priori scales. The four essentialism items, for example, were averaged to form a scale. Analysis, therefore, began with a check of inter-item correlations, factor structure, and scale reliability in both pre- and posttests. Although some controversy attaches to the issue of degrees of freedom in statistical analyses of group-level manipulations (Kenny, Kashy, & Bolger, 1998), in this paper we take the perspective that the responses of individual participants can be treated as independent observations. Accordingly, data across the three seminars were pooled.

After the checks on scale structure and reliability, posttest scores of pretested participants were compared to posttest scores of non-pretested participants to check for pretest sensitization effects. In the absence of pretest sensitization, a within-subjects comparison of the pre- and posttest scores provided by participants who had both was performed to test for seminar effects. The within-subjects approach has greater power, but in cases where the pretest did seem to have an effect, the influence of the seminar was assessed using a between-subjects comparison of pretest scores and the posttest scores of posttest-only participants.

Demographic Measures

The following 11 demographic items were assessed: age, gender, sexual orientation, disability (if any), race, first-learned language, most preferred language, religion, ethnic/cultural heritage, parental income, and preferred South African political party. The order of demographic items remained constant, although the order of nonsequential response options (e.g., languages, such as SiSwati, Afrikaans, and English) was re-randomized between pre- and posttests.

Subjective Evaluations of the Seminar

Perceived seminar effectiveness. There were five items for assessing participants’ estimations of the effectiveness of the seminar. In the pretest, they were phrased as requests for estimates of how effective participants expected the seminar to be. The first effectiveness item asked how successful the participant expected the seminar to be (or in the posttest, how successful it was) in teaching the skill of effective interethnic interaction to the average student. The item was rated on a 7-point scale ranging from 1 (not successful) to 7 (very successful).
The second and third items asked how much the average student would learn, on a 7-point scale ranging from 1 (none) to 7 (much); and how much the average student’s behavior would change as a result of their participation, on a 7-point scale ranging from 1 (no change) to 7 (great change). The fourth and fifth items asked the latter two questions, but about the participant’s own learning and behavior. In the posttest, the two items concerning the respondent’s self were augmented by open-ended adaptations: “In what ways, if any, will you now think or speak differently than you did before the workshop about members of other groups?” and “In what ways, if any, will you now act differently than you have in the past towards members of other groups?” High levels of omitted or uninterpretable responses, however, made it impossible to learn much from the open-ended augmentations.

The five posttest seminar effectiveness items that had pretest counterparts as measures of expected seminar effectiveness were averaged to form a posttest scale of perceived seminar effectiveness with a Cronbach’s alpha of .88. The pretest items were averaged to form a pretest scale of expected seminar effectiveness ($\alpha = .87$).

The variance of perceived seminar effectiveness scores was different for pretested ($SD = 0.92, n = 46$) and non-pretested ($SD = 1.34, n = 35$) participants, $t(79) = 4.26, p = .042$. A separate-variance $t$ test was undertaken, therefore, to compare the mean ratings given by the two groups. The test indicates that perceived seminar effectiveness was higher among pretested ($M = 5.83$) than among non-pretested ($M = 5.12$) participants, $t(57) = 2.94, p = .005$ (two-tailed), $\eta^2 = .108$. The comparison of pretested ($M = 4.97, SD = 1.11, n = 47$; see Table 1) and non-pretested participants’ expected seminar effectiveness ($M = 5.12, SD = 1.34$) indicated no reliable difference, $t(80) = 0.35, p = .726$ (two-tailed) $\eta^2 = .002$. A two-tailed $t$ test indicates that participants who had been through a similar seminar before rated the current program’s effectiveness as lower ($M = 5.07, SD = 1.31, n = 20$) than did participants who had not been through such a seminar ($M = 5.72, SD = 1.04, n = 55$), $t(73) = 2.02, p = .047, \eta^2 = .053$.

**Rankings of seminar features.** One of the “Final Items” in the posttest asked participants to rank 15 different aspects of the seminar they had just experienced (Table 2). They were requested to rank each aspect according to which were “the most important or meaningful.” In the revised posttest, for U2 and U3, the feature “The History Line exercise” was replaced by “Learning what might help me avoid being stereotyped by others.”

Despite a request to rank the 15 seminar features (including a write-in feature labeled “Other”) using each rank number once and only once, participants often assigned duplicate ranks or none at all. It also was common to rank a number of the most important or meaningful features with the
highest ranks (i.e., lowest rank numbers), a number of the least meaningful with the lowest ranks, and to leave blanks in the middle. To further complicate the picture, a typographical error in the item presented to U2 and U3 participants indicated that there were 16 features on the list, when in fact there were only 15.

Using all of the ordinal information provided by each participant, their raw ranks were replaced with a consistent, complete set of ranks from 1 to 15. Duplicated ranks were replaced by the average rank of the positions occupied by the duplicates. For example, four ranks of “1” became four ranks of 2.5, and the next item was awarded a rank of 5, even if the next raw

Table 1

Pretest and Posttest Means and Standard Deviations

<table>
<thead>
<tr>
<th>Scale</th>
<th>n</th>
<th>Range</th>
<th>Pretest M</th>
<th>SD</th>
<th>Posttest M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceptions of the workshop</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seminar effectiveness(^\text{a})</td>
<td>47, 35</td>
<td>1-7</td>
<td>5.0</td>
<td>1.1</td>
<td>5.1</td>
<td>1.3</td>
</tr>
<tr>
<td>Negative reactions(^\text{b})</td>
<td>79</td>
<td>1-3</td>
<td>1.9</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall evaluation(^\text{b})</td>
<td>83</td>
<td>1-3</td>
<td>2.9</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Seriousness of stereotyping</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ease of making mistakes</td>
<td>47</td>
<td>1-7</td>
<td>4.6</td>
<td>1.5</td>
<td>5.2(^*)</td>
<td>1.5</td>
</tr>
<tr>
<td>Race-based treatment of self</td>
<td>46</td>
<td>1-7</td>
<td>4.2</td>
<td>1.6</td>
<td>4.3</td>
<td>1.5</td>
</tr>
<tr>
<td>Race-based suffering</td>
<td>45</td>
<td>1-7</td>
<td>6.0</td>
<td>1.5</td>
<td>6.5(^*)</td>
<td>0.7</td>
</tr>
<tr>
<td><strong>Interracial opinions and attitudes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimize ethnic differences</td>
<td>44</td>
<td>1-7</td>
<td>4.9</td>
<td>2.0</td>
<td>4.5</td>
<td>2.0</td>
</tr>
<tr>
<td>Attend less to ethnic differences</td>
<td>45</td>
<td>1-7</td>
<td>4.1</td>
<td>2.2</td>
<td>4.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Support ethnic differences</td>
<td>45</td>
<td>1-7</td>
<td>5.3</td>
<td>1.7</td>
<td>5.7</td>
<td>1.6</td>
</tr>
<tr>
<td>Differences between races</td>
<td>46</td>
<td>1-7</td>
<td>4.6</td>
<td>1.1</td>
<td>4.5</td>
<td>1.2</td>
</tr>
<tr>
<td>Essentialism</td>
<td>45</td>
<td>1-7</td>
<td>4.8</td>
<td>1.4</td>
<td>5.2</td>
<td>1.3</td>
</tr>
</tbody>
</table>

**Note.** Means and standard deviations are rounded to one decimal place.

\(^{\text{a}}\)Posttest scores of pretested participants were not equivalent to those of non-pretested participants on this scale, suggesting an effect of the pretest itself. Seminar effectiveness was therefore assessed using a between-subjects comparison of pretest scores (\(n = 47\), “expected effectiveness”) scores and posttest-only scores (\(n = 35\) “perceived effectiveness”). Seminar impact for other scores and scales used a within-subjects comparison of pretest and posttest scores for 47 pretested participants.

\(^{\text{b}}\)Administered at posttest only. *Significantly different from pretest value at \(p < .05\).
rank given by the participant was “2.” Blanks were considered to be ties for the positions not assigned explicitly by the participant, except in one case. If the lowest assigned rank (highest number) was one less than the number of items the instructions stated there were (i.e., 14 for U1; 15 for U2 and U3) and the rank of “Other” as well as that of at least one additional feature were left blank, then it was assumed that “Other” was intended to have the lowest rank permitted by the instructions (i.e., 15 for U1; 16 for U2 and U3). The remaining blank features were given tied ranks for the unassigned positions.

The ranks assigned to highly and poorly ranked features were piled, of necessity, into one end or the other of the spectrum of possible ranks and, therefore, were skewed. Consequently, the median rank assigned to each

Table 2

Participants’ Median Ranks of Seminar Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hearing other people’s stories</td>
<td>5.1</td>
</tr>
<tr>
<td>2. Understanding what stereotypes are and how they work</td>
<td>5.3</td>
</tr>
<tr>
<td>3. Addressing misperceptions/disagreements through direct dialogue with others</td>
<td>5.6</td>
</tr>
<tr>
<td>4. Learning more about what the stereotypes of particular groups are</td>
<td>6.3</td>
</tr>
<tr>
<td>5. Developing clear steps to take to reduce reliance on stereotypes</td>
<td>6.8</td>
</tr>
<tr>
<td>6. Learning what might help me avoid being stereotyped by others</td>
<td>6.8(^a)</td>
</tr>
<tr>
<td>7. Experiencing what it feels like to be stereotyped</td>
<td>6.9</td>
</tr>
<tr>
<td>8. The facilitators</td>
<td>7.2</td>
</tr>
<tr>
<td>9. Simply interacting with members of other groups</td>
<td>7.6</td>
</tr>
<tr>
<td>10. Identifying important features of your own identity</td>
<td>7.9</td>
</tr>
<tr>
<td>11. History Line exercise</td>
<td>8.0(^b)</td>
</tr>
<tr>
<td>12. Information Circle exercise</td>
<td>9.2</td>
</tr>
<tr>
<td>13. Telling your own story</td>
<td>9.4</td>
</tr>
<tr>
<td>14. The venue</td>
<td>12.5</td>
</tr>
<tr>
<td>15. The dance party</td>
<td>12.9</td>
</tr>
<tr>
<td>16. Other</td>
<td>14.7</td>
</tr>
</tbody>
</table>

\(^a\)This item was not on the initial list of items, which was presented to participants from U1. The reported median rank is based on its consideration by participants from U2 (\(n = 35\)) and U3 (\(n = 18\)). \(^b\)This item was only presented on the list of items to be ranked by participants from U1 (\(n = 26\)).
feature of the seminar was taken as a summative indicator of its appeal to participants. The 15 medians from each seminar were distributed normally, and the correlations of those from one workshop with those from another were very high, with all one-tailed $ps < .01$ ($r = .84$ for U2 and U3; $r = .78$ for U2 and U1; $r = .72$ for U1 and U3). Supported by these strong correlations, medians computed over all ($n = 79$) participants for each feature appear in Table 2. Note that the four most popular features include interpersonal communication (Features 1 and 3) and learning more about stereotypes (Features 2, 4, and 5).

**Best thing about the seminar.** Presented immediately after the feature-ranking task, this open-ended item asked “If not indicated above, what was the most significant or best thing about the workshop for you?” Responses to the item fell into two categories: communication and situation. The situational references were in the minority and indicated appreciation of such things as the food, the facilitators, or going for a hike. Most responses fell into the communication category and referred to the opportunity to interact across group boundaries in general and to hear the personal stories of others in particular. Examples include “dialogue opportunity with all parties present,” “spending an extended period of time, hearing what people had to say,” and “the willingness of all to participate.” The responses evoked a clear sense that delegates experienced a type of relief at being able to communicate about historically divisive and awkward issues in a safe environment.

Notes taken through participant observation during the first seminar confirmed the value that delegates placed on interaction. Several people told their personal stories to the entire group, and one of these was especially dramatic. It was the story of how the narrator’s father had been politically active during the struggle against apartheid and was beaten by the police right in front of the narrator, who was only 6 years old or so at the time. The father was taken away and never returned. When the remaining participants told their own stories in small groups, people listened intently, though only a few of the stories were as dramatic as the first one. The sharing of the narratives, which were sometimes quite personal, seemed much appreciated. Almost without exception, the stories went beyond the allotted 15 min per person. The participants clearly got to know one another better and seemed to develop a higher degree of empathy for and solidarity with the others than they had had before.

After the individual stories, participants interacted across race groups in plenary. At one point, the assembly was to consider the possibility that there might be a grain of truth to some stereotypical traits and to think about how they might have come about. This task sparked some of the liveliest debate of the entire program. Though the facilitators noted later that the reaction of the U1 participants in this exercise were not typical, certain U1
participants eventually sidestepped completely the premise that there might be some understandable source of the stereotypical traits associated with them. Instead, they insisted on having it explained to them how anyone could have come to associate such words with them in the complete absence of any justification, however remote or exaggerated.

**Possible adverse reactions to the seminar.** Three items were devised to assess possible negative reactions to the program. They were answered by circling one of three possible responses: *No* (1), *Not sure* (2), or *Yes* (3). Appropriate only to the posttests, the items were as follows: “Can you think of anyone in the seminar who might have been hurt (seemed overly discouraged, depressed, or embarrassed)?”; “Can you think of anyone in the seminar who became angry when her/his views were challenged?”; and “Can you think of anyone in the seminar who might have become more prejudiced as a result of their participation?” Each item was followed by the parenthetical instruction to “include yourself, if this question is appropriate to you.”

The three posttest side-effects items averaged to form a scale with a Cronbach’s alpha of .67, a respectable result, considering that there were only three items and that each item had only three response options. The overall mean negative reaction score was 1.91 (SD = 0.59, n = 79; Table 1). Participants with previous seminar experience perceived more negative seminar effects \( (M = 2.24, \ SD = 0.61, \ n = 19) \) than did first-time seminar participants \( (M = 1.81, \ SD = 0.63, \ n = 56) \), \( t(73) = 2.43, \ p < .017 \) (two-tailed), \( \eta^2 = .075 \).

**Overall evaluation of the seminar.** The summative question “On balance, do you think the seminar was worth the time and resources put into it?” appeared as the final item on the posttest questionnaire. It was answered on the 3-point (*No*/*Not sure*/*Yes*) scale.

A great proportion of the participants (i.e., 79 of 83 responding, or 94%; Table 1) agreed after the seminar that the program had been worth the time and resources expended on it. Of the remaining four participants, two were unsure, and only the other two indicated that the seminar was not worthwhile.

**Recommendations for change.** The penultimate item was the open-ended question “In your opinion, how should the workshop be changed in order to improve the program in the most important way possible?” Responses to this item fit sensibly into four categories: content, process, participants, and miscellaneous.

1. **Content of seminar.** A notable number of suggestions were made that prejudice and stereotyping based on other social groupings (e.g., gender, class) be considered in the seminar.

2. **Process of seminar.** There were clear indications that delegates felt they would benefit from more time to process and debrief both
informational and emotional aspects of the seminar. This indicated that delegates experienced a sense of either being “left behind” regarding the theory and principles of stereotyping or of being left with unresolved emotional issues. The prevalence of suggestions in this regard indicates that facilitators need to be more aware and considerate of “where the delegates are at” and perhaps to allow more time for debriefing, even at the expense of program completion.

3. Participants. Suggestions in this category included trying for more diverse groups and also for equal representation of the groups that are present. Delegates were aware of the issue and perceived it as a weakness in the process.

4. Miscellaneous. These responses included abstractions, generalizations, and rhetoric regarding stereotyping and prejudice in society in general, and did not provide a useful source of information.

Information About Stereotyping

Five items were included on the questionnaire to assess the program’s effectiveness at communicating particular information about stereotyping and about the potential hazards of stereotypical thinking. The first two were multiple-choice items on the definition and function of stereotypes, and the last three were attitudinal items assaying the perceived seriousness of the overuse of stereotypes.

Definition and function of stereotypes. A multiple-choice question was used to assess participants’ understanding of what a stereotype is: “A stereotype is ________ someone because of the group to which they belong.” Participants were instructed to circle the letter in front of the phrase that most correctly fills the blank. The correct answer was “information presumed true about.” The three foils were “an uncomplimentary opinion about,” “unfair treatment of,” and “All of the phrases are correct.” On the pretest, 27 of the 47 respondents (57.5%) were able to correctly identify the social psychological definition of a stereotype. On the posttest, 21 respondents (44.7%) were correct on this item (there was no significant difference by sign test).

The stereotype function question was also a multiple-choice question that read as follows: “People use stereotypes ...” Participants were instructed to circle the letter in front of the most correct completion of the sentence, with the following options: “to speak badly about others without being blamed,” “in order to take unfair advantage of others,” “because stereotypes are correct,” and “because stereotypes simplify the world’s
complexity.” The correct option was the lattermost. The number of participants correctly recognizing the function of stereotypes was not different from pretest to posttest (pretest: 24 of 47 participants, 51.1%; posttest: 29 of 47 participants, 61.7%; not significant by sign test).

**Seriousness of stereotype-related problems.** Two of the three items on the gravity of stereotype-related problems were judgment scales, and the third was an opinion item. The scale items were “How easy is it to make mistakes when interacting with people from another culture?” which was rated on a 7-point scale ranging from 1 (not at all easy) to 7 (extremely easy); and “In your dealings with other people, how often do you think that the way they treat you is based on your race?” which was rated on a 7-point scale ranging from 1 (virtually never) to 7 (virtually always). The opinion item was “Many South Africans still suffer tremendously because others perceive them in terms of their racial group.”

After the seminar, the 47 pre- and posttested participants rated it easier to make mistakes when interacting with people from another culture (M = 5.23, SD = 1.49) than they did before (M = 4.61, SD = 1.51; Table 1), \( t(46) = 2.21, p = .016, \eta^2 = 0.096 \). There was no evidence of a change (pretest, M = 4.19, SD = 1.61; posttest, M = 4.30, SD = 1.52) in participants’ opinions (n = 46; Table 1) of how often they were treated as they were because of their race, \( t(45) = 0.25, p = .400, \eta^2 = .001 \). Participants’ (n = 45) agreement that South Africans still suffer tremendously because of the racially oriented perceptions of others increased significantly over the course of the seminar, \( t(44) = 2.66, p = .006 \) (one-tailed), \( \eta^2 = .138 \), from a mean of 6.01 (after converting the original −3-to +3 scale to a 1-to-7 scale; SD = 1.49) to a mean of 6.50 (SD = 0.71; Table 1).

**Effects on Attitudes and Opinions**

Measures of participants’ professed opinions are presented roughly in order of the level of the scale at which their topic may be placed. Items at a more specific, interpersonal level, such as the number of a participant’s close friends from different racial backgrounds, appear toward the end; items at a more global, societal level, such as participants’ opinions about affirmative action and their ideal for the relationships between ethnic groups, follow immediately.

**Models of ethnic relations.** Three opinion items were designed to assess the attractiveness of different, hypothetically independent philosophies regarding interethnic relations. Conceptually, the first of these world views was focused on an actively homogenizing, melting-pot goal that implied the establishment of a single culture throughout the population (i.e., “Ethnic
differences should be minimized in order to emphasize the common humanity that links people of all cultures”).

The guiding principle of the second opinion item was the passive, tolerant, mosaic society ideal in which cultural differences need neither to be blended into a common purée, nor purposely reinforced (i.e., “Ethnic differences should get less attention; these differences often distract us from improving health, education, and jobs for all”). The third philosophy was based on the rainbow-nation model, in which cultural differences are deliberately sought after and celebrated (i.e., “Ethnic differences make our country stronger and more interesting; these differences should be recognized and supported”).

Separate tests of seminar effects, all of which were nondirectional, were conducted for each ethnic model item; none of these tests produced a significant result. Pretest agreement with the proposition that ethnic differences should be minimized (\(M = 4.91, SD = 2.01, n = 44\)) was similar to posttest agreement (\(M = 4.52, SD = 2.01; \text{Table 1}\)), \(t(43) = -1.23, p = .227, \eta^2 = .034\). Agreement with the idea that ethnic differences should get less attention was the same on the pretest (\(M = 4.10, SD = 2.24, n = 45\)) as it was on the posttest (\(M = 4.11, SD = 2.21; \text{Table 1}\)), \(t(44) = -0.07, p = .944, \eta^2 = .000\). Finally, agreement that ethnic differences should be supported was unchanged from pretest (\(M = 5.33, SD = 1.72, n = 45\)) to posttest (\(M = 5.71, SD = 1.54; \text{Table 1}\)), \(t(44) = 1.38, p = .176, \eta^2 = .041\).

Regarding affirmative action. The two assertions pertaining to affirmative action were as follows: “Even today in South Africa, people’s racial backgrounds exert a huge impact on their opportunities and welfare,” and “People in formerly disadvantaged racial categories must be given special consideration in hiring, education, and promotion decisions, even when their credentials on paper are not as good as their competition’s.” There was a distinct ceiling effect on participants’ agreement with the first item, regarding opportunities and welfare, resulting in a very high mean score (\(M = 6.14, \text{on a scale ranging from 1 to 7}\)) and an even higher median (\(Mdn = 6.3\)).

In contrast, agreement with the proposition that some people must be given special consideration despite weaker credentials was platykurtic, \(g(2) = -1.37, \text{SE}[g(2)] = 0.53\), and possessed of two distinct modes (at \(x = 2\) and \(x = 5\); \(Mdn = 2.8, M = 3.44\)). Respondents either disagreed with the proposition (\(n = 46\)) or agreed with it (\(n = 32\)), and only a few (\(n = 3\)) could not say. The two variables were not significantly correlated (Spearman’s \(r_S = .16, ns\)). Tested using Wilcoxon’s signed-ranks procedure, differences between pre- and posttest responses to neither the impact item, \(T = 99.5, ns\); nor the special-consideration item, \(T = 93.0, ns\), could be considered reliable.
Perceived differences between and within races. The first judgment scale item was “How great are the cultural differences between the following groups?” There followed a list of all six possible pairings of the four race groups in which the order of each group’s appearance was determined randomly. Each pair of groups was accompanied by a 7-point scale ranging from 1 (not great) to 7 (very great). The second judgment scale item was “How great is the variety of abilities, opinions, and experience WITHIN each of the following groups?” The four race groups were listed individually, each accompanied by the same scale as in the last item.

There was a strong expectation that inter- and intra-race heterogeneity ratings, racial identification scores, and levels of comfort with the prospect of approaching strangers of different races would be related to the participant’s own race group. Given that these expectations were realized, seminar effects would manifest not simply by changing the level of the variable in question, but by changing the way it interacted with participant race. In other words, the seminar was hypothesized to reduce the advantage enjoyed by the participant’s own race group over other race groups on these variables. Therefore, the Participant Race H Target Race interaction, instead of the simple effect of the variable itself, was checked and then, if present, used in the corresponding analysis.

When ratings of inter- and intra-race heterogeneity were examined for the expected advantage of participant race, ANOVA results indicated no such effect. Neither participant race, target race, nor their interaction had a significant effect on inter- or intra-race heterogeneity ratings.

The consensus among participants regarding subjective racial heterogeneity ratings implied that these ratings could be combined into scales. Factor analyses of the six inter-race items and the four intra-race items suggested one factor for each set of items. That is, one factor represented the extent to which a respondent felt that there were great cultural differences between racial groups, whichever groups were being compared, and the other the extent to which a respondent felt that there was a great variety within groups, again irrespective of the particular group in question. Scales formed by finding the average of the ratings for each factor displayed good reliability (intergroup, \( \alpha = .79 \); intragroup, \( \alpha = .88 \)) and were moderately negatively correlated (\( r = -.24, p = .034 \)).

The mean inter-race difference ratings of the pretested participants (\( n = 46 \)) were much the same after the seminar (\( M = 4.48, SD = 1.17 \)) as they were before (\( M = 4.61, SD = 1.08 \); Table 1), \( t(45) = -0.53, p = .298 \) (one-tailed), \( \eta^2 = .006 \). The pretested participants (\( n = 45 \)) also rated the variety within races much the same after the seminar (\( M = 5.13, SD = 1.44 \)) as they did before (\( M = 5.04, SD = 1.11 \); Table 1), \( t(44) = 0.19, p = .426 \) (one-tailed), \( \eta^2 = .001 \).
Essentialism. The idea that different groups have inherently different natures or essences can be a powerful barrier to intergroup understanding (Gil-White, 2001). To assess essentialist thinking, the following four opinion items were used: (a) “If people from different ethnic backgrounds were raised the same, they would have similar interests and abilities”; (b) “If people from different ethnic backgrounds were raised the same, they would have similar occupations and incomes”; (c) “If people from different ethnic backgrounds were raised the same, they would still differ in musical and artistic expression”; and (d) “If people from different ethnic backgrounds were raised the same, they would still differ in personality style.”

The two positively worded essentialism items were reverse-scored and averaged with the two negatively worded items to produce pre- and posttest essentialism scales, each with a Cronbach’s alpha of .64. The posttest essentialism scores of the 45 people who completed both pre- and posttest items were not lower ($M = 5.24, SD = 1.31$) than their pretest scores ($M = 4.78, SD = 1.43$; Table 1), $t(44) = 2.07, p = .978$ (one-tailed), $\eta^2 = .088$.

Note the trend toward a higher posttest mean. Had the hypothesis predicted an increase or a nondirectional difference, rather than a decrease, this result would have been significant. In the end, the whole group of seminar participants ended up with a substantial essentialism score ($M = 5.03, SD = 1.33, n = 79$).

Stereotypical trait lists. Without the posttest lists, potential changes induced by the seminar, if any, could not be detected. The pretest lists provided a modicum of qualitative information, to wit: Listed traits of racial groups were overwhelmingly pejorative. They also were generally consistent from respondent to respondent. No tendency for individuals to provide more positive traits for any one group (thus indirectly identifying their own racial in-group) was apparent.

Social identification with different racial groups. The stem for the first identification question was “How different do you feel to the average person at your university who is …?” which was followed by a list of the four race groups. Participants were instructed to rate their responses on a 7-point scale ranging from 1 (not different) to 7 (very different). The second stem was “How strong are the ties you feel to others at your university who are …?” It was followed by a fresh list of race names and scales, identical to the last list except that the names were in a different order.

The two items for assessing identification with each of the four race groups appeared on both pretest and posttest, yielding four pairs of identification items on the pretest and four on the posttest. The correlations between items in a pair ranged from $- .26$ (posttest identification to the Coloured group) to $- .60$ (pretest identification with the Black group). On the basis of these correlations, the rating of perceived difference in each pair
was reversed and averaged with its mate, the rated strength of ties, to produce an index of identification. There were thus eight of these indexes, each based on two items.

Because participants were expected to identify more strongly with their own race group than with others, a 3 (Respondent Race: Coloured, White, or Black) × 4 (Identification Race: Coloured, White, Black, or Indian) × 2 (Seminar: pretest vs. posttest) repeated-measures ANOVA of identification, with identification race and seminar as within-subjects factors, was used to test effects of the seminar. As expected, there was a significant interaction between the race of the respondent and the race of identification, $F(6, 123) = 23.35$, $p < .001$, $\eta^2 = .533$. The bias toward own-race identification, however, was not altered by the seminar, as the three-way Race × Identification Race × Seminar interaction was not significant, $F(6, 123) = 0.96$, $p = .454$, $\eta^2 = .045$ (see Table 3 for mean identification scores, taken over pre- and posttests).

The main effect of identification race was significant, $F(3, 123) = 24.48$, $p < .001$, $\eta^2 = .374$, indicating that, as a whole, the participants included in this analysis—who were all either Black, White, or Coloured (the 3 Indian respondents were too few to be included)—identified less strongly with Indian university colleagues than with others. None of the other effects in the analysis were significant.

**Shortfall of extra-racial close friends.** In the context of four other self-relevant posttest “Final Items” (see Perceived seminar effectiveness), participants were asked to provide estimates of how many close friends they had and how many cross-race close friends they had. The difference between the

<table>
<thead>
<tr>
<th>Respondent race</th>
<th>Coloured</th>
<th>White</th>
<th>Black</th>
<th>Indian</th>
<th>Over all others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coloured</td>
<td>6.4</td>
<td>0.6</td>
<td>4.3</td>
<td>1.1</td>
<td>4.9</td>
</tr>
<tr>
<td>White</td>
<td>4.5</td>
<td>0.7</td>
<td>5.7</td>
<td>0.7</td>
<td>4.0</td>
</tr>
<tr>
<td>Black</td>
<td>4.5</td>
<td>1.0</td>
<td>3.9</td>
<td>1.2</td>
<td>6.1</td>
</tr>
<tr>
<td>Cross-race mean</td>
<td>4.9</td>
<td>1.2</td>
<td>4.6</td>
<td>1.3</td>
<td>5.1</td>
</tr>
</tbody>
</table>

*Note.* Ratings were made on a 1–7 scale. Means taken over both pretests and posttests. Means and standard deviations are rounded to one decimal place.
two was taken as an index of how uninvolved a delegate was cross-racially. In both the pretest and the posttest, most people reported a modest number of close friends (of either type), but a few reported great numbers (e.g., 138 close friends and 74 from other races). To normalize the resulting skewed distribution, the difference scores (e.g., 64 more in-group than out-group close friends) were transformed by taking their base-10 logarithm. Some people had zero differences, though, so 0.5 was added to each difference before the logarithm was taken to prevent the log (of 0) from being returned as undefined. After the seminar, the pretested participants (n = 39) scored no lower on the uninvolvment index (M = 0.67, SD = 0.43; Table 1), t(38) = −0.80, p = .816 (one-tailed), η² = .006.

Projected comfort approaching strangers. A series of four judgment scales appeared for assessing respondents’ comfort levels with approaching strangers from their own and other races. The questions read “If you were taking your lunch to sit down at a table of ______________ students you didn’t know, how comfortable about it would you feel?” In each item, one of the race-group names appeared in lieu of the blank, and responses were rated on a 7-point scale ranging from 1 (not comfortable) to 7 (very comfortable).

A 3 (Participant Race: Coloured, White, or Black) × 4 (Stranger Race: Coloured, White, Black, or Indian) × 2 (Seminar: pretest vs. posttest), repeated-measures ANOVA of comfort levels, with stranger race and seminar as within-subjects factors, reveals a significant Participant Race × Stranger Race interaction, F(6, 114) = 11.02, p < .001, η² = .367, indicating that respondents were more comfortable when approaching unfamiliar students of their own race than those of other races (see Table 4). This own-race preference was not different after the seminar from what it was before (i.e., the three-way interaction was not significant), F(6, 114) = 0.96, p = .454, η² = .048. A significant main effect of the seminar, F(1, 38) = 4.70, p = .037, η² = .110, however, indicates that the seminar increased the average level of comfort reported by participants from every group (Black, White, and Coloured) for approaching unknown students from every group (pretest and posttest Participant × Target Group means are not tabled). The resulting change in the grand mean was from a pretest mean of 4.50 (SD = 1.19) to a posttest mean of 4.92 (SD = 1.24).

The main effect of stranger race was also significant, F(3, 114) = 4.99, p = .003, η² = .116, highlighting the relatively low ratings of comfort with approaching Indian strangers given by the Black, White, and Coloured participants. The Stranger Race × Seminar and the Participant Race H Seminar interactions were not significant—F(3, 114) = 0.98, p = .961,

---

6The three Indian respondents were too few to be included in this analysis.
Table 4

Mean Projected Comfort While Approaching Strangers

<table>
<thead>
<tr>
<th>Stranger race</th>
<th>Coloured</th>
<th>White</th>
<th>Black</th>
<th>Indian</th>
<th>Over all strangers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respondent race</td>
<td>$n$</td>
<td>$M$</td>
<td>$SD$</td>
<td>$n$</td>
<td>$M$</td>
</tr>
<tr>
<td>Coloured</td>
<td>8</td>
<td>6.2</td>
<td>1.0</td>
<td>15</td>
<td>4.5</td>
</tr>
<tr>
<td>White</td>
<td>15</td>
<td>4.5</td>
<td>1.0</td>
<td>15</td>
<td>5.5</td>
</tr>
<tr>
<td>Black</td>
<td>18</td>
<td>4.6</td>
<td>1.7</td>
<td>18</td>
<td>4.2</td>
</tr>
<tr>
<td>Mean</td>
<td>41</td>
<td>4.9</td>
<td>1.5</td>
<td>4.6</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Note. Ratings were made on a 1–7 scale. Means were taken over pretests and posttests. Means and standard deviations are rounded to one decimal place.


\[ \eta^2 = .003; \text{ and } F(2, 38) = 0.29, \ p = .747, \ \eta^2 = .015, \] respectively—indicating that the disadvantage of Indian strangers did not change over the course of the seminar and that respondents of all three examined racial groups reacted to the seminar similarly. Finally, there was no detectable main effect of participant race, $F(2, 71) = 0.18, \ p = .836, \ \eta^2 = .009$, providing no evidence of any difference in the overall comfort levels of the three tested participant groups.

Discussion

Overall, the current evaluation indicates little effect of the seminar program on participant knowledge or beliefs. As described more fully in the next section, however, it does indicate some affective change. These findings are based on quantitative, before-and-after comparisons of participants’ attitudes and beliefs, complemented by qualitative records of participants’ explicit, generally quite favorable, evaluations of the seminars and their results. Thus, the discussion begins with a review of pretest–posttest assessments of program product, then moves to a consideration of participant commentary regarding both program process and product. Included is a mention of those remarks by participants and the direct observations by researchers that suggest a possible basis for seminar’s appeal to the people who took part.

Seminar Effects on Attitudes and Beliefs

By and large, there was no evidence that the seminars affected interethnic attitudes. Two such measures, comfort with approaching cross-race
strangers and identification with own race group, are particularly telling because they revealed distinct in-group favoritism in both pre- and posttests. Similarly, there was no decrease in the shortfall of extra-race close friends, and though measures of inter- and intra-race heterogeneity were not associated with the participant’s own race group, those perceptions of heterogeneity were not discernibly affected by the seminar. More global attitudes regarding the ideal for interethnic relations and affirmative action also were unchanged. Finally, participants’ understanding of what stereotypes are and what they do did not improve.

The test for a reduction in participants’ tendency to essentialize racial differences provided no evidence of such an effect. Furthermore, it came as a considerable surprise that the posttest scores were not only no lower than pretest scores, but apparently higher instead. If it were true that there had been no effect of the seminars at all on essentialism (and if our pretest data accurately represent the population of possible scores), then there would be only 1 chance in 50 ($p = .02$) to obtain as great an increase in essentialism as the one we observed. This result provides a compelling motivation to re-examine the issue. Recourse to the growing body of literature on the phenomenon of essentialism (e.g., Gil-White, 2001) may provide insight into the effect observed here. If so, it may suggest future adjustment either to the hypothetical predictions of seminar effects, to seminar procedures, or both.

The seminar did affect two beliefs about the negative consequences of stereotyping. One was the belief that a great deal of suffering results from stereotyping; participants saw more such suffering after the seminar than before it. Second was the belief that it is easy to make mistakes when interacting with those from other races; participants saw such mistakes as easier after participating in the seminar. Investigation of these effects awaits further research, but it does seem clear that the program, with its pivotal experiential exercise of reading out each group’s stereotypical trait words to them as they stand to hear them, reiterated to participants the aversive character of prejudicial treatment.

Participants’ Evaluation of the Seminars

Participants’ posttest ratings of seminar effectiveness were generally very high (averaging 5 on a unipolar, 7-point scale) and unchanged from pretest ratings of expected seminar effectiveness. Participants with previous experience in such programs were less positive, an unexpected result that requires investigation in future research. Participants’ open-ended responses identified no particular seminar effect on either cognition or behavioral intention, nor any particular hurt to seminar participants. Most notable is the fact that
94% of the participants asserted that the program was worth the not-in-
considerable resources of time and money that were invested in it.

The feature rankings proved to be a rich source of information regarding
participant opinions. Four of the five top-ranked aspects of the program
(Table 2) involved learning the views of others or addressing discrepancies
between those views and the participants’ own. Participants’ appreciation of
dialogue was reiterated in their open-ended descriptions of what they found
to be best in the seminar, and direct observation of group interaction con-
Firmed that the brief autobiographies provided in the last Saturday session
riveted the attention of listeners in each small group. Moreover, the first
Sunday session, in which participants were encouraged to identify prejudices
based on inaccuracy or ignorance, generated the most animated conversa-
tion of the program. According to participant observations made at the U1
seminar, incongruities between in-group and out-group views of people’s
own groups were of central importance.

Thus, it appears that, in the participants’ eyes, the best aspect of the
seminar was the opportunity it afforded to discover, or confirm, the per-
spectives on their own group held by those from other groups, with whom
they generally may be expected to have had relatively little personal inter-
action; and when discrepancies between those perspectives and their con-
structions of their own identities were identified, to address the discrepancies
through dialogue. The importance of this opportunity and its potential im-
pact were not anticipated and were not evaluated adequately in this study.
Although a dialogue on discrepant constructions of social identities may be
thought of as a kind of intergroup education, it is debatable whether it is an
education participants would seek of their own accord. A deeper consid-
eration of this issue in future work seems advisable.

Pettigrew (1998) asserted that intergroup contact is likely to have positive
effects only if it comes with friendship potential, that is, the possibility for
sustained, appealing social interaction. Consistent with this suggestion, and
with Neville and Furlong’s (1994) findings, the present results reinforce the
conclusion that short-term programs, without sufficient prospect for future
interaction or follow-up, are not likely to have much effect on the beliefs,
attitudes, or behavioral predispositions of participants. The limited impact
of program instruction about the nature of stereotypes indicates that even
the understanding of stereotyping—let alone the application of stereotypes
—can be surprisingly resistant to change.

Note that the impact of short-term programs is apt to be limited despite
the explicit, highly positive evaluations of seminar participants. To some
extent, participant reactions may be subject to demand effects (Orne, 1969).
Respondents motivated to help and to reciprocate the efforts of workshop
facilitators are likely to try to provide the kind of results the facilitators
appear to be seeking. In addition, participants in prejudice-reduction activities are likely to be those who are already in favor of such programs’ explicit, egalitarian objectives (Neville & Furlong, 1994). Thus, participants may evaluate positively a program of unproven effectiveness simply to encourage the continued pursuit of a goal they view as laudable. On the other hand, participants may be sensitive to genuine benefits, such as the opportunity to address perceived discrepancies between the image they hold of their group and the images of their group held by others, which are not being detected well by current measures. The basis of the positive participant reactions requires further investigation.

The possibility that contact-based programs may have unexpected effects must also be borne in mind. In our results, the startling indications regarding increased essentialist thinking must be followed up without delay. Also, an affective reaction to unjust treatment based on an arbitrary demographic classification was aroused, and it seems to have affected participants’ perceptions of the level of suffering such treatment engenders. That this kind of affective arousal might motivate constructive reform of intergroup behavior, belief, and regard seems plausible, but it might also, on occasion, stimulate less constructive behaviors not anticipated by intervention planners.

Perhaps the clearest indication of the beneficial potential of adequately supported programs similar to the ones under study here can be found in the increased comfort participants felt after the seminar with respect to encountering strangers from a variety of racial backgrounds, including their own. It seems that the opportunity for earnest, reasonably in-depth dialogue can reassure participants about their capacity for meeting, understanding, and interacting with others, a capacity that is clearly of interest to many who find themselves in increasingly diverse societies. If confirmed with behavioral reports from longer term studies, development of this kind of capacity might be a major attraction for contact-based programs. Many educational institutions, from preschools to universities, advertise the educational value of a diverse student body, but there is some doubt about how much interaction actually occurs across cultural boundaries, even with institutional support (McCauley, Plummer, Moskalenko, & Mordkoff, 2001). Contact-based programs may do more of what student diversity promises. Diversity offers the potential for interaction, but program intervention may be necessary, in many cases, to move potential into reality.

When applied to programs meant to improve relations between ethnic and other groups in business, government, or the voluntary sector, our observations of limited impact suggest that expectations regarding the results of short-term programs be moderated and that sweeping claims of program efficacy be considered critically by organizational and public policymakers. Conversely, a commitment to achieve marked, sustained effect
on attitudes and behaviors is likely to require a considerable, sustained investment of time and material resources. Participant experiences, which occupy a key theoretical position in the conduct of the programs, ought to be made as appealing and as future-oriented as possible. Beyond this kind of success, there are bound to be surprises waiting for researchers and program planners who begin evaluating the diversity initiatives and other contact interventions undertaken to improve relations between divided groups.

In cases in which an adequate investment of resources eventually is made, care must be taken to monitor for and to avoid counterproductive effects while accentuating the desirable ones. This is especially important when efforts are made to recruit program participants who might not be committed already to the values and approaches of most diversity initiatives. When participant and program objectives are not consonant, the need to monitor program effects and to make the appropriate adjustments is increased. Of course, there are ethical, as well as practical concerns surrounding the recruitment of and planning for participants who might harbor reservations. Within the bounds of ethical responsibility, the central concern must be to ensure that available resources are directed toward those techniques and programs that best contribute to genuine and lasting intergroup conciliation.

References


Hewstone, M., Cairns, E., Voci, A., Paolini, S., McLernon, F., Crisp, R. J., et al. (2005). Intergroup contact in a divided society: Challenging seg-


