Changing Heterosexuals' Attitudes Toward Homosexuals: A Systematic Review of the Empirical Literature

Edmon W. Tucker and Miriam Potocky-Tripodi

*Research on Social Work Practice* 2006; 16; 176
DOI: 10.1177/1049731505281385

The online version of this article can be found at:
http://rsw.sagepub.com/cgi/content/abstract/16/2/176
Changing Heterosexuals’ Attitudes Toward Homosexuals: A Systematic Review of the Empirical Literature

Edmon W. Tucker
Miriam Potocky-Tripodi
Florida International University

Objective: This article systematically reviews evidence for interventions that change attitudes toward homosexuals.

Method: In all, 17 empirical studies using college and/or university student samples and interventions intended to improve heterosexuals’ attitudes toward lesbian, gay, or bisexual individuals are reviewed. Characteristics of the studies with regard to participants, interventions, methodologies, attitude measures, and outcomes are reported. The studies are also rated for their level of empirical support. Finally, challenges of the research and applications to social work practice are discussed.

Results: No intervention met the criteria of a well established or probably efficacious treatment, as all studies had substantial methodological limitations.

Conclusion: These interventions require further testing with well designed, methodologically sound experiments to determine efficacy. Researchers who pursue this controversial topic, however, may experience considerable opposition in obtaining funding and/or forums for dissemination of their findings.

Keywords: attitudes; homosexual; heterosexual; gay/lesbian/bisexual; systematic literature review

Antigay attitudes in the form of heterosexism and homophobia are pervasive (Yang, 1997) and create significant sources of stress and/or pain for those in the sexual minority (Herek, 1992). The ideological system of heterosexism, which “denies, denigrates, and stigmatizes any non-heterosexual form of behavior, identity, relationship, or community” (Herek, 1990, pp. 316-317), has been promulgated by societal institutions in this country, such as the courts, religion, medicine, and the mass media. This heterosexism has historical roots that suggest that homosexuality is perceived as a threat to Western society (Fone, 2000).

Although American society remains divided on lesbian, gay, and bisexual (LGB) issues, including whether antigay attitudes should be changed, the official position of the profession of social work is unambiguous. Social workers are prohibited from discriminating against clients or colleagues based on their sexual orientation (National Association of Social Workers, 1999). The social work code of ethics also states that social workers should work to “prevent and eliminate domination of, exploitation of, and discrimination against any person . . . on the basis of . . . sexual orientation” (National Association of Social Workers, 1999). The National Association of Social Workers has released various public policy statements concerning sexual orientation, including that “same-gender sexual orientation should be afforded the same respect and rights as other-gender orientation” (National Association of Social Workers, 2002). Furthermore, schools of social work should be teaching their students to practice without discrimination and “with respect, knowledge and skills related to clients’ . . . sexual orientation” (Council on Social Work Education, 2001).

It is not surprising that antigay attitudes are highly correlated with antigay behaviors, including physical attacks (Franklin, 2000; Patel, Long, McCammon & Wuensch, 1995; Roderick, McCammon, Long, & Allred, 1998; Whitley, 2001). In fact, alarming numbers of LGB people are subjected to discrimination, harassment, and violence because of their sexual orientation (Berrill, 1992; D’Augelli & Grossman, 2001; D’Augelli, Pilkington, & Hershberger, 2002; Herek, Gillis, Cogan, &Glunt, 1997; Rose & Mechanic, 2002; Thurlow, 2001; Waldo, Hesson-McInnis, & D’Augelli, 1998). The exact extent of this victimization cannot be known for certain.

Official reports of hate crimes represent only a small fraction of the actual number of incidents motivated by the victim’s sexual orientation. Many of these crimes go unreported, and when reported, there are problems with
the collection and assimilation of the data (Federal Bureau of Investigation, n.d.; Kuehnle & Sullivan, 2003; Rose & Mechanic, 2002). Moreover, LGB victims often do not notify law enforcement, because they expect an unsympathetic or even hostile response from the police. This expectancy may be based on either their own prior experiences with law enforcement personnel, the shared experiences of others, or both (Berrill & Herek, 1992; Herek, Gillis, & Cogan, 1999).

We do know, however, that the empirical research reveals that homosexuals are much more likely than heterosexuals to be the victims of violent crimes (Berrill, 1992; Bontempo & D’Augelli, 2002; DuRant, Krowchuk, & Sinal, 1998; Herek, Gillis, Cogan, & Glunt, 1997). Similarly, LGB individuals endure far more discrimination and harassment than their heterosexual peers (Bontempo & D’Augelli, 2002; D’Augelli, 1992; Lewis, Derlega, Berndt, Morris, & Rose, 2001; Thurlow, 2001). Pervasive antigay attitudes also mean that the sexual minority is often denied equal access to housing, employment and/or promotions, education, and health care (Hunter, Joslin, & McGowan, 2004). Furthermore, same-sex couples in the United States are refused the fundamental right to marry and are, therefore, deprived of all the social, legal, and financial benefits that marriage conveys to heterosexual couples.

The psychological consequences and other attendant effects of this victimization and discrimination should not be overlooked. Given the correlation between antigay attitudes and behaviors, this systematic review was conducted to determine what, if any, empirically validated interventions exist for improving heterosexuals’ attitudes toward homosexuals. This review entailed analysis of selected studies in relation to their methodological characteristics and findings.

**METHOD**

**Selection Criteria**

Studies for inclusion in the analyses were identified through searches of the PsycINFO, Social Services Abstracts, and Sociological Abstracts electronic databases from January 1994 to August 2004. Only studies that included at least one intervention designed to improve the heterosexual participants’ attitudes toward homosexuals and that were published in peer-reviewed journals were included. Search parameters were broad and included wild cards. Specifically, the search terms were **attitude** and **homosexual** or **gay** or **lesbian** and **experiment** or **intervention** or **outcome** or **change**.

As particular attitudes are often culture bound (Stycos, 1998; Evans, 1997), we limited our inquiry to studies conducted within the United States. Furthermore, a study must have assessed the participants’ personal attitudes toward homosexuals or homosexuality generally to be included. Thus, a study that used participant opinions regarding a specific, politically controversial issue (e.g., gays in the military, gay marriage, etc.) to assess attitudes would not suffice for purposes of this analysis. Only one study was eliminated for this reason.

**Rating Criteria**

Level of empirical support was assessed using criteria developed by the American Psychological Association’s Division 12 task force for evaluating empirically validated therapies (Chambless et al., 1998; Chambless & Hollon, 1998). None of the studies reviewed herein used single-case designs; therefore, only the criteria for group designs were used. Only interventions that have been demonstrated to be superior (statistically significantly so) to a placebo or another intervention or to be equivalent to an already established treatment in at least two good, between-group design experiments with adequate sample sizes (at least 25 participants per condition) are considered well established (Chambless et al., 1998).

A good between-group design means that participants were randomly assigned to the intervention of interest or to one or more comparison conditions (i.e., randomized, clinical trials; Chambless & Hollon, 1998). Additionally, the experiments must have been conducted with treatment manuals, unless the intervention was relatively simple and adequately specified in the procedure section of the journal article reporting on its efficacy. Furthermore, the characteristics of the participants must have been clearly specified, and the experiments must have been conducted by at least two different investigators or investigating teams. If an intervention met the criteria above, except it was demonstrated to be superior in only one experiment, or it was demonstrated to be superior in two or more experiments conducted by the same investigating team, it was considered probably efficacious. Alternatively, if two or more experiments demonstrated that an intervention was superior to a waiting-list control group, it was considered probably efficacious if the experiments met all of the other criteria.

There are other important elements of establishing efficacy. Demonstrating superiority assumes that the outcome assessment tools have demonstrated reliability and
validity (Chambless & Hollon, 1998). Multiple methods of assessment are preferable, although not required, and participant self-reports are suspect. Evaluators are also cautioned to check that researchers have interpreted their outcome data correctly. In other words, assessing outcomes and interpreting data appropriately are critical components of good between-group design experiments (Chambless & Hollon, 1998, p. 8).

A study rating sheet was created in accordance with the criteria outlined above. A random sample of six of the eligible studies was selected to test for interrater reliability between the first and second authors. With a reliability of 83.3% established, the first author rated the remaining studies for empirical soundness.

FINDINGS

Table 1 summarizes the 17 studies that met our selection criteria. This section presents the pertinent findings and summarizes the commonalities across studies with regard to participants, interventions, methodologies, attitude measures, and outcomes. Finally, the studies are evaluated for their level of empirical support.

Summary of the Studies’ Participants

Interventions are often efficacious for only a specific problem or population; therefore, it is important that investigators adequately describe any characteristics of the participants that might affect the generalizability of their findings (Chambless & Hollon, 1998). Specifically, it is the participants who completed all aspects of the experiment and about whom postintervention data are available who should be described rather than some larger prescreening pool. In many cases, only the larger pool of eligible participants is described; therefore, attrition, dropout, and/or missing data information should be explained so the reader can make a judgment about whether the final sample is representative of the described participants. Because more than half (nine) of the studies here under review either did not adequately address attrition (four studies) or experienced so much attrition (attrition rates greater than 10% or five studies), we cannot be confident that the final respondents were adequately described in those articles. All of the reviewed studies used convenience samples comprised of either undergraduate or graduate students in U.S. schools. The remainder of our comments about the participants, however, will focus on the eight studies where the final respondents were adequately described.

All of the studies presently under review reported participants’ gender. Overall, there were wide disparities in gender representation. Of the eight studies where participants were adequately described, four had large (60% or more) majorities of females (Black, Oles, Cramer, & Bennett, 1999; Cotton-Huston & Waite, 2000; Guth, Lopez, Clements, & Rojas, 2001; Probst, 2003). Two of the studies had less than a 10% difference in gender disparity among participants (Corley & Pollack, 1996; Grutzeck & Gidycz, 1997), and two had male majorities (Grack & Richman, 1996; Wallick, Cambre, & Townsend, 1995).

Of the eight participant-described studies, six addressed the ages of the participants by providing either a range, median, or mean. Given that all participants were pursuing either undergraduate or graduate degrees, the age range was limited. With few exceptions, the participants were between the ages of 18 and 35.

Five of the eight participant-described studies addressed participants’ ethnicity. One of these (Guth et al., 2001) reported only the percentage of Caucasians, which, as with all of the studies, was the overwhelming majority. The remaining four studies reported percentages of participants in five ethnic categories that included Caucasian or White, African American or Black, Hispanic, Asian American or Pacific Islander, Native American, and Other (Black et al., 1999; Grutzeck & Gidycz, 1997; Probst, 2003; Wallick et al., 1995).

With regard to other participant characteristics, all of the eight participant-described studies reported the type of undergraduate or graduate courses from which the respondents were recruited. Six of the eight reported participants’ sexual orientation. Two reported marital status. One reported on participants’ contact with homosexuals, and another included whether participants had gay or lesbian friends. Finally, one reported on participants’ income, religion, region, and parental acceptance of lesbians and gay men, and one reported number of participants with a disability.

Summary of the Studies’ Interventions

The vast majority of the 17 articles devoted substantial attention to the theoretical bases of their investigations. Most focused on two main justifications for their studies: (a) a cognitive and/or educational function of the intervention (e.g., to dispel myths and stereotypes attributed to homosexuals), and (b) contact theory—that exposure to and shared positive experiences with homosexuals will help to change heterosexuals’ prejudices. In fact, only

(text continues on p. 184)
TABLE 1: Summary of Studies’ Characteristics

<table>
<thead>
<tr>
<th>Authors</th>
<th>Participants</th>
<th>Interventions</th>
<th>Methodologies</th>
<th>Attitude Measures</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bassett &amp; Day, 2003</td>
<td>48 of 65 first-year MSW students enrolled in three sections of a required course. Demographics based on the pool of 65. % Female: 87.7 Mean age: 34 Ethnicity: 76.9% White 13.8% Black 6.2% Hispanic 3.1% Asian American</td>
<td>The required human behavior course was infused with content (both written and audiovisual) about and regarding gay men. No comparison. No control.</td>
<td>Pre- and posttest given at the beginning and end of the semester. Attrition reported and attributed primarily to absences, withdrawals, and confusion regarding self-assigned numeric codes used to ensure anonymity.</td>
<td>Modified version of Attitudes Toward Lesbians and Gay Men (ATLG; Herek, 1988). Modifications were specified. Validity and reliability not reported.</td>
<td>Authors reported that the total group of participants had significantly lowered mean score on the ATLG posttest.</td>
</tr>
<tr>
<td>Black, Oles, Cramer, &amp; Bennett, 1999</td>
<td>56 social work students. % Female: 86 Mean age: 29 Ethnicity: 78.6% White 14.3% Black 5.4% Hispanic 1.8% Asian American</td>
<td>90-minute speaker panel of 2 gay men and 2 lesbians (a university administrator, a tenured professor, and 2 social work students), who discussed various issues related to their sexual orientation and answered questions throughout. No comparison. No control.</td>
<td>Pre- and posttest given 1 week prior to and 1 week following presentation. Attrition reported.</td>
<td>Modified version of ATLG. Modifications were specified. Reliability reported.</td>
<td>Authors reported no significant change in participants’ scores from pre- to posttest.</td>
</tr>
<tr>
<td>Corley &amp; Pollack, 1996</td>
<td>141 heterosexual Introduction to Psychology students. % Female: 55 Mean age: 19.6 (female) 19.26 (male) Ethnicity: Not reported. Participants per group not reported, but there were 12 groups for analyses (E1, E2, and E3 divided by gender and by traditional vs. nontraditional).</td>
<td>E1: Nonstereotypical written description of lesbian couple (feminine-feminine) E2: Stereotypical written description of lesbian couple (masculine-feminine) E3: Somewhat stereotypical written description of lesbian couple (masculine-masculine) No control.</td>
<td>Random assignment. Posttest immediately following intervention. Follow-up 1 week later. Attrition reported.</td>
<td>Heterosexual Attitudes Toward Homosexuality Scale (HATH; adapted to assess for lesbians only; Larsen et al., 1980). Validity reported via citation. Reliability reported.</td>
<td>Traditional and nontraditional females who received E1 were significantly more positive toward lesbians than those who received E2, but differences were not sustained in follow-up. For traditional males, significant differences existed between all three conditions and were sustained. For nontraditional males, there were no significant differences. Insufficient power to support analyses.</td>
</tr>
<tr>
<td>Cotton-Huston &amp; Waite, 2000</td>
<td>150 heterosexual undergraduate business and psychology students. % Female: 83 Age: 70% were 18-22 Ethnicity: Not reported E1: 53 E2: 50 C1: 47</td>
<td>E1: 45-minute video depicting homosexual lifestyles and celebrating gay pride E2: classroom interaction with a gay panel during one class period C1: No intervention.</td>
<td>Posttest only. Attrition reported.</td>
<td>Index of Attitudes Toward Homosexuals (IAH; Hudson &amp; Ricketts, 1980). Reliability reported.</td>
<td>No significant differences between groups on attitudes reported.</td>
</tr>
</tbody>
</table>
### TABLE 1: Summary of Studies' Characteristics

<table>
<thead>
<tr>
<th>Authors</th>
<th>Participants</th>
<th>Interventions</th>
<th>Methodologies</th>
<th>Attitude Measures</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cramer, 1997</td>
<td>107 MSW students in six foundation practice sessions.</td>
<td>E1: Instructor self-disclosed her lesbian status while conducting educational unit focusing on lesbian identity development.</td>
<td>Pre- and posttest. Posttest given 6 weeks after intervention. Attrition reported. Baseline comparisons not reported.</td>
<td>Modified version of ATLG. Modifications somewhat specified. Reliability reported for instrument and subscales (Attitudes Toward Lesbians [ATL] and Attitudes Toward Gay Men [ATG]) but not for modified versions.</td>
<td>Students in E1 had significantly lower mean score (more positive attitudes) than those in C1 on the ATG at posttest.</td>
</tr>
<tr>
<td>Cramer, Oles, &amp; Black, 1997</td>
<td>Started with 110 undergraduate and graduate social work students at four different schools.</td>
<td>Four different educational interventions, using the information-plus-exposure model and varying disclosure or nondisclosure of instructor sexual orientation, given during the course of one semester. No control.</td>
<td>Pretest given during first 2 weeks of classes. Posttest given during last 2 weeks of classes. Amount of attrition not reported, but missing data were excluded from analyses.</td>
<td>Modified version of ATLG. Modifications somewhat specified. Reliability reported for instrument and subscales (ATL and ATG) but not for modified version.</td>
<td>There was significant change in all four classes from pretest to posttest on ATLG scores. Comparison between the four educational approaches is problematic, as the instructors varied on several dimensions. No significant differences between groups were found. Sufficiency of power for analyses questionable.</td>
</tr>
<tr>
<td>Finkel, Storaasli, Bandele, &amp; Schafer, 2003</td>
<td>48 graduate students at University of Denver's Graduate School of Professional Psychology.</td>
<td>Two 2-hour “Safe Zone” diversity training sessions separated by 6 months. Treatment manuals mentioned. No comparison. No control.</td>
<td>Posttest at end of second session that required participants to (a) rate current level of homophobia and (b) retrospectively rate level of preintervention homophobia. Attrition reported.</td>
<td>Riddle Homophobia Scale (Wall, 1995). Authors reported that the psychometric properties of the scale were unknown.</td>
<td>No significant differences between current and retrospective ratings.</td>
</tr>
<tr>
<td>Grack &amp; Richman, 1996</td>
<td>37 undergraduate psychology students who had scored above the 50th percentile on the Gay and Lesbian Attitude Scale (GLAS; Grack &amp; Richman, 1996).</td>
<td>All groups instructed to solve series of logic problems. E1: 2 actors self-identifying as homosexual participants and group received extra credit reward E2: 2 actors self-identifying as homosexual participants and group did not receive reward E3: 2 actors self-identifying as heterosexual participants and group rewarded E4: 2 actors self-identifying as heterosexual participants and group did not receive reward. No control.</td>
<td>Pretest given 8 weeks prior to the 1-hr intervention. Posttest immediately following intervention. Attrition not reported. Baseline comparisons not reported.</td>
<td>GLAS, which was reportedly based on Hudson and Ricketts' (1980) Index of Homophobia Scale. Reliability for the Index of Homophobia Scale was reported but authors do not explain how their scale differs from the original, nor is the validity or reliability of the GLAS itself reported.</td>
<td>There was a significant difference reported between groups based on the sexual orientation of the actors. Insufficient power to support analyses.</td>
</tr>
</tbody>
</table>
Grutzeck & Gidycz, 1997

200 undergraduate introductory psychology students from a moderately sized midwestern university.

% Female: 54
Age: Not reported
Ethnicity:
85% White
11% Black
1% Asian American
1% Native American
2% Other
E1: 63
E2: 64
C1: 73

E1: 1-hour speaker panel presentation of 4 undergraduates (2 gay males, 2 lesbians), who gave brief biographical sketches followed by question-and-answer period.

E2: Participants were given a handout with basic facts and harmful stereotypes associated with homosexuality.

C1: Participants were told that more than expected had shown up for the experiment, so it was not necessary for them to participate but that they would still receive credit.

Random assignment. Pretest given 2 weeks prior to intervention. Posttest given 4 weeks after intervention. Attrition not reported.

Modified version of the IAH (modifications not specified) and the HATH. Psychometric properties of both instruments and of modified IAH were reported. Behavioral measure developed by the authors (psychometric properties indeterminate).

Panel presentation did not have a significant effect on either instrument; however, the entire sample had more tolerant scores on the IAH at posttest. Likely pretest effect. Behavioral measure outcomes indeterminate.

Guth, Lopez, Clements, & Rojas, 2001

47 upper-level undergraduate psychology students at a southeastern university.

% Female: 78.7
Median age: 29.5
Ethnicity: 88% White
E1: 17
E2: 15
C1: 15

Both E1 and E2 were 2-hour workshops that contained myths and facts about homosexuality and everyday issues faced by gays and lesbians.

E1: Content was presented "rationally"
E2: Content was presented "experientially"

C1: Workshop that focused on finding psychology-related material on the Internet. Guidelines for interventions mentioned.

Random assignment. Pre- and posttest. Posttest given 3 weeks after intervention. Attrition reported.

Thought-listing procedure developed by authors that involved having participants spontaneously list thoughts elicited by a stimulus (six hypothetical scenarios dealing with homosexuals). Three raters (all with doctorates in psychology) judged the pre- and posttests to create categories, then a panel of three graduate students categorized the thoughts. Interrater reliability was reported. Validity not reported.

Thought categories changed significantly at posttest only for E2 group. No between-group comparisons. Insufficient power to support analyses.

Hood, Muller, & Seitz, 2001

150 undergraduate students in five sections of an organizational behavior course.

% Female: 49.5
Mean age: 24.8
Ethnicity:
55.7% White
28.1% Hispanic
6.8% Asian American
1% Black
1% Native American
3.5% Other

Course content promoting diversity competency.

Vague description of intervention. No comparison. No control.

Pretest given at the beginning of semester. Posttest given during one of the last two classes. Attrition reported.

A five-item Attitude Toward Gay Men and Lesbians Scale developed by the authors. Coefficient alphas for the pre- and postintervention surveys were reported. No other psychometric properties reported.

Change for entire sample was reportedly significant. No significant change by racial group.

(continued)
### TABLE 1: Summary of Studies' Characteristics

<table>
<thead>
<tr>
<th>Authors</th>
<th>Participants</th>
<th>Interventions</th>
<th>Methodologies</th>
<th>Attitude Measures</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson &amp; Krieger, 1997</td>
<td>190 psychology students from six separate classes surveyed across three semesters at a mid-sized southeastern university.</td>
<td>50-minute panel presentation by 2 gay male and 2 lesbian students attending the same university, who gave personal narratives followed by question-and-answer period. No comparison.</td>
<td>Pretest given approximately 2 weeks prior to intervention. Posttest given 6 weeks after intervention. Attirion not reported.</td>
<td>Modified version of the Attitudes Toward Homosexuality Scale (ATHS; MacDonald &amp; Games, 1974). Modifications somewhat specified. Psychometric properties not reported.</td>
<td>Change in sample's posttest score reported as significant. Females demonstrated more change in attitude than males.</td>
</tr>
<tr>
<td>Pratarelli &amp; Donaldson, 1997</td>
<td>80 heterosexual students at a large midwestern university.</td>
<td>E1: Written educational scenario supporting a biological explanation for homosexuality</td>
<td>Random assignment. Baseline survey given 1 week prior to intervention. Posttest immediately following intervention. Attrition not reported.</td>
<td>Two matched, prenormed surveys designed to assess changes in participants' attitudes toward homosexuals, developed by first author. Reliability reported.</td>
<td>No statistically significant differences between groups were reported.</td>
</tr>
<tr>
<td>Probst, 2003</td>
<td>57 undergraduate students in a workplace diversity psychology course at Washington State University made up the experimental group</td>
<td>E1: A semester long (17 weeks) upper-level undergraduate workplace diversity psychology course that included content on gay, lesbian, and bisexual issues. C1: An elementary statistics course taught by the same instructor during the same semester.</td>
<td>Pre- and posttest given during the first and last weeks of the semester, respectively. No significant differences between groups on pretest.</td>
<td>Homonegativity Scale (six items; Morrison, Parriag, &amp; Morrison, 1999). Reliability reported.</td>
<td>There was a significant improvement in the attitudes of E1 at posttest, whereas C1 attitudes were slightly worse (more negative), suggesting that a between-group comparison would be significant. No between-group comparisons.</td>
</tr>
<tr>
<td>Riggle, Ellis, &amp; Crawford, 1996</td>
<td>72 students in introductory psychology courses.</td>
<td>E1: An 88-minute documentary, &quot;The Times of Harvey Milk,&quot; about one of the first openly gay elected officials in the United States C1: Completed the posttest prior to viewing the documentary</td>
<td>Pretest 3 weeks prior to intervention. Posttest immediately following. Attrition reported and notable (from 314 down to 72). Between-group differences at prescreening reported (not significant).</td>
<td>ATHS developed by Herek, 1984. Citation given for psychometric properties.</td>
<td>E1 had significantly less prejudicial attitudes at posttest.</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Intervention</td>
<td>Outcome</td>
<td>Notes</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>Waldo &amp; Kemp, 1997</td>
<td>156 undergraduate students in introductory psychology course</td>
<td>E1: In one section of the course, the instructor came out (self-identified as gay) to the class midway through the semester while presenting educational unit about sexual orientation; C1: Four sections of the same course that were taught by self-identified heterosexual instructors and covered same material regarding sexual orientation.</td>
<td>Pre- and posttest at the beginning and end of the semester, respectively. Attrition not reported. Baselines reported.</td>
<td>E1 scores changed significantly at posttest compared to controls.</td>
<td></td>
</tr>
<tr>
<td>Wallick, Cambre, &amp; Townsend, 1995</td>
<td>Three consecutive 1st-year classes of medical students at Louisiana State University in New Orleans (exact number of participants provided only for the class entering 1991).</td>
<td>3-hour panel presentation by 3 gay and lesbian physicians, who provided autobiographical sketches, and a faculty member, who shared his son's coming out story, followed by a question-and-answer period; then, students broke into smaller groups for discussion.</td>
<td>Of the 186 students in the 1991-1992 cohort, 180 completed the IAH at the beginning of their freshman year, 168 at midyear (2 weeks following intervention), 114 at year's end, and 185 following their required psychology clerkship in their 3rd year. Participation rates were reportedly similar in the two subsequent years, but the IAH was given on entrance and following the intervention only.</td>
<td>An overall 6.3% decrease in homophobic attitude within the pooled (all three classes) data was reported (p &lt; .03). On follow-up for the '91 class, there was a rebound effect following the junior clerkship.</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** E = experimental group condition; C = control group condition.

a. The Index of Homophobia Scale is the same instrument cited by other authors in this study as the IAH. The creators of this instrument suggested the name change to IAH to reduce potential bias in response to the original name (Hudson & Ricketts, 1980).

b. Riggle et al. either incorrectly identified Herek as the developer of the ATHS or correctly cited to Herek but called the scale by the wrong name. The name of the instrument developed by Herek (1984, 1988) is the ATLG. An instrument entitled the ATHS was developed by MacDonald and Games (1974).
three of the studies did not rely to some extent on one or both of these rationales (Pratarelli & Donaldson, 1997; Probst, 2003; Wallick et al., 1995).

Pratarelli and Donaldson (1997) were short on theory but they were investigating whether a biological explanation for sexual orientation influenced attitudes toward homosexuals. Probst (2003) made reference to the increase in diversity courses on college campuses and to the paucity of empirical assessment regarding the extent to which such courses actually change students’ attitudes and behaviors. Wallick et al. (1995) saved their rationale for the conclusion, where they simply cited a policy paper adopted by the American Medical Association calling for physicians to demonstrate a nonjudgmental attitude toward gay men and lesbians and for medical schools to increase their focus on how to appropriately address the needs of this population.

As for the interventions themselves, 11 of the studies infused course content with or held workshops presenting educational information about homosexuality. Formats varied and included verbal, written, and audiovisual communications. Typically, the stigmatization and discrimination experienced by LGB people were addressed, as were common myths and stereotypes. Most of the interventions took place within the time span of one class period.

Five of the studies used gay and/or lesbian speaker panels as interventions. Three involved the coming out (disclosure of homosexual status) of the classroom instructor to the student participants in combination with an educational unit about sexual orientation. Other interventions included a nonstereotypical written description of a lesbian couple, a logic problem-solving exercise in which two actor participants self-identified as homosexuals to the other participants, and a written biological explanation for sexual orientation.

As previously mentioned, with the exception of relatively simple interventions that have been adequately explained in the journal article reporting their efficacy, treatment manuals have been deemed essential to empirically validated interventions (Chambless & Hollon, 1998). Without them, other researchers cannot know precisely what treatment was tested, nor can the intervention be replicated. Only 2 of the 17 reviewed articles mentioned the use of treatment manuals or their equivalents (Finkel, Storaasli, Bandele, & Schafer, 2003; Guth et al., 2001).

Summary of the Studies’ Methodologies

Only 4 of the 17 studies reviewed were true experimental designs in which participants were randomly assigned to intervention, comparison, and/or control conditions. Unfortunately, 3 of the experiments also used a pretest. Compounding this problem, none of the 3 used a design to account for any possible pretest effects.

Thirteen of the 17 studies were quasi-experimental designs that did not randomly assign participants to comparison groups. In fact, in 6 of the 13, there were no comparison or control groups, which precluded the possibility of between-group comparisons. In 5 of these no comparison–no control studies, there was simply one group of participants who were given one or more pretests, interventions, and posttests. The remaining study simply used a posttest that asked participants to rate their current level of homophobia and to also retrospectively rate their level of homophobia prior to the intervention (Finkel et al., 2003). Of the 7 quasi-experiments that used comparison and/or control groups, 6 of them conducted pretests, but only 3 of them discussed the issue of baseline differences between groups. Of these, only 1 had reported a low level of attrition that did not compromise the equivalency of the groups (Probst, 2003).

Summary of the Studies’ Attitude Measures

It should be noted that many of the 17 reviewed studies were assessing participants on multiple variables. For example, 1 study assessed participants’ masculinity or femininity, attitudes regarding women’s rights and roles, demographics, and attitudes toward homosexuals (Cotton-Huston & Waite, 2000). Our discussion is limited to those instruments that measured participants’ attitudes toward homosexuals.

Six of the 17 studies used measures of heterosexuals’ attitudes toward homosexuals developed by other investigators in previous research. Two used an index developed by Hudson and Ricketts (Cotton-Huston & Waite, 2000; Wallick et al., 1995), and 1 used a scale developed by Herek (Riggle, Ellis, & Crawford, 1996). Both of these measures have established validity and reliability. One investigating team, however, used the Riddle Homophobia Scale developed by Wall (1995; Finkel et al., 2003). This instrument simply asks participants to concurrently (postintervention) rate their current level of homophobia on an 8-point Likert-type scale and rate what their level of homophobia had been prior to the intervention using the same scale. These authors reported that the psychometric properties of the Riddle Homophobia Scale were
unknown but that they deemed it to have acceptable face validity.

One of the studies used the Homonegativity Scale developed by Morrison, Parriag, and Morrison (1999; Probst, 2003). This measure contains six items using a 5-point Likert-type scale. The investigator reported the test-retest reliability, but we would like to have seen some discussion regarding the instrument’s validity. In this regard, however, a reference for the instrument’s psychometric properties was cited in the study.

Grutzeck and Gidycz (1997) were the only investigators to use multiple measures to assess participants’ attitudes toward homosexuals. They implemented the Heterosexual Attitudes Toward Homosexuality Scale (HATH) developed by Larsen, Reed, and Hoffman (1980) and a modified version of Hudson and Ricketts’ Index of Attitudes Toward Homosexuals (IAH; Hudson & Ricketts, 1980), both of which have previously established validity and reliability. The modifications to the IAH were not specified, but the psychometric properties of the modified version were reported. Grutzeck and Gidycz also used a behavioral measure that they developed for the study to assess tolerance for homosexuals. We have serious doubts about the validity of the behavioral measure, the limitations of which are clearly articulated by the investigators themselves.

Nine studies used modified or adapted versions of previously established measures. These studies varied regarding the extent to which the modifications were specified and whether the psychometric properties of the modified versions were reported. Most, however, were vague about the modifications and did not report on their properties. Of the 14 studies that used either full or modified (or both, i.e., Grutzeck & Gidycz, 1997) versions of preestablished instruments, 6 utilized measures originally developed by Herek and 4 used some version of Hudson and Ricketts’s (1980) IAH. The only other instrument used or adapted by more than one of the studies was Larsen et al.’s (1980) HATH, used by 2 of the studies.

Finally, three of the studies used an instrument developed by the investigators of those studies as the sole measure of attitudes toward homosexuals (Guth et al., 2001; Hood, Muller, & Seitz, 2001; Pratarelli & Donaldson, 1997). Guth et al. (2001) developed an instrument using a thought-listing technique for their pretest and posttest measures. They provided a rationale for this type of measurement, supporting reference citations, and documentation of the interrater reliability. Absent, however, was any discussion of validity for this particular instrument.

Hood et al. (2001) developed their own scale because they were interested in attitudes toward gay men and lesbians specifically with regard to workplace issues. This instrument was composed of five questions on a 5-point Likert-type scale. Coefficient alphas for the pretest and posttest scores were provided, but again, there was no discussion of the validity of this scale.

Pratarelli and Donaldson (1997) used “two matched prenorned surveys” (p.1412) previously developed by the first author with another investigator to assess changes in participants’ attitudes. Two reference citations presumably regarding the development of this instrument were provided, but one referenced a paper presented at a convention, and the other referenced a manuscript submitted for publication. In the present article, reliability was reported, but there was no discussion of the validity of this instrument.

Summary of the Studies’ Analyses and Outcomes

The methodological limitations of the studies summarized above renders a discussion of their analyses and outcomes practically moot. That is, statistical significance is irrelevant when a model has been misspecified or an assumption has been violated. Statistical computer programs are not able to correct for research design and/or methodological shortcomings (Pedhazur & Schmelkin, 1991). There were, however, additional limitations concerning the studies’ analyses that are worth mentioning.

Of the nine studies that analyzed differences between experimental and comparison or control groups, only four used either random assignment without a pretest (one) or nonrandomly assigned groups with baseline comparisons on pretests reported (three). Of these four, only one (Waldo & Kemp, 1997) had enough participants per group to power the analyses. This study, however, did not report the amount of attrition.

None of the reviewed studies discussed effect size or clinical significance. Reports of statistical significance alone are of limited utility (Chambless & Hollon, 1998; Pedhazur & Schmelkin, 1991). We hope that as this area of study expands, there will emerge meaningful units of measures for antigay attitude change. Ultimately, scholars will want to know how much of a change in attitude is required for a measurable change in behavior.

Evaluation of the Level of Empirical Support of the Interventions

None of the interventions qualified as well established, as no single intervention was subjected to two independent experiments (Chambless et al., 1998; Chambless &
Hollon, 1998). Nor did any of the interventions meet the criteria for probably efficacious treatments. Most fell short at this level because they were not tested in a between-group design experiment.

Random assignment serves to equate the comparison groups on all variables except for the intended manipulations. Without it, many researchers believe that the comparison groups are inherently and immutably nonequivalent (Chambless & Hollon, 1998; Pedhazur & Schmelkin, 1991). Nevertheless, investigators will employ various methods (e.g., establishing no significant differences between groups on a pretest) in an attempt to accomplish valid comparisons.

As previously mentioned, four of the reviewed studies used random assignment (Corley & Pollack, 1996; Grutzeck & Gidycz, 1997; Guth et al., 2001; Pratarelli & Donaldson, 1997). Two of these used outcome measures that were developed by one or more of the investigators, and the validity of these instruments were not sufficiently established (Guth et al., 2001; Pratarelli & Donaldson, 1997). Additionally, one of these did not conduct between-group comparisons, and there was insufficient power to support such analyses (Guth et al., 2001). The other one did not report participant attrition, nor did it control for possible pretest effects (Pratarelli & Donaldson, 1997).

Of the two randomized studies that used measures with previously established validity and reliability, one did not have sufficient power to conduct its analyses (Corley & Pollack, 1996). The other did not report using treatment manuals for the experimental group that was exposed to a speaker panel presentation. Additionally, this study did not report attrition, nor did it control for possible pretest effects, and the results indicated that such effects were in operation (Grutzeck & Gidycz, 1997).

**DISCUSSION AND APPLICATIONS TO SOCIAL WORK RESEARCH AND PRACTICE**

This evaluation for empirical validation revealed that none of the interventions were adequately tested by these studies. Optimum design and methodology is an expensive proposition, sometimes prohibitively so. Therefore, investigators can hardly be faulted for making due with their limited resources.

That said, some of the problems reported in our findings warrant further discussion. With regard to participant characteristics, the investigators properly reported that all respondents were university or college students. At a minimum, we thought that investigators should also report on participants’ gender, age, ethnicity, and religiosity and/or religion.

Prior research has shown that there are significant differences in attitudes and behaviors toward homosexuals across gender (Franklin, 2000; Whitley, 2001; Yang, 1997). Logic dictates that the outcomes of interventions designed to change attitudes might also differ by gender. Therefore, it is imperative that investigators conducting such studies report on the gender makeup of their samples and any significant differences in outcomes between male and female participants.

Like gender, age could be a distinguishing participant characteristic. From a developmental perspective, we know that attitudes can evolve throughout the life span (Pillari, 1998). Specifically, surveys have demonstrated that there are generational disparities with respect to attitudes toward homosexuals (Ricci & Biederman, 2004). Therefore, it would not be surprising if intervention effects varied by age group.

There is little empirical evidence one way or the other regarding the effect of ethnicity on attitudes toward homosexuals. Ethnicity has, however, had a demonstrated effect on outcomes for therapeutic interventions (Arroyo, Miller, & Tongan, 2003; Markowitz, Spielman, Sullivan, & Fishman, 2000). Therefore, even if there were not enough participants of various ethnicities in the present studies to demonstrate between-group effects, we would still want to know their ethnic makeup so that we can make a judgment about the generalizability of the findings.

Antigay attitudes have been consistently correlated with religion and religiosity (e.g., Hinrichs & Rosenberg, 2002; Laythe, Finkel, Bringle, & Kirkpatrick, 2002; Snively, Kreuger, Stretch, Watt, & Chadha, 2004). Specifically, studies have found that the participants who score high on measures of antigay attitudes also tend to be those who are the most religious. Similarly, those who belong to fundamentalist churches score higher, overall, on measures of homophobia and/or heterosexism. Therefore, it is important to know how intervention effects vary by religion and religiosity. It is notable that only four of the seventeen reviewed studies attempted to investigate these effects (Black et al., 1999; Cotton-Huston & Waite, 2000; Cramer, 1997; Cramer, Oles, & Black, 1997).

The investigators of at least one study (Waldo & Kemp, 1997) purposely did not request demographic information other than gender from participants in an effort to ensure anonymity and to reduce demand effects. Anonymity can be ensured, however, through study design elements while still collecting valuable demographic
information. Participants should be instructed that the surveys are designed to be anonymous and to avoid writing their names or other identifying information on their answer sheets. Additionally, research assistants unknown to the participants could administer the surveys and instruct the respondents that the assistant will not be personally analyzing the data. A slotted, locked box could be placed in the room where participants deposit their surveys when completed. Furthermore, it is best if investigators can obtain participants with whom they have no other contact (e.g., as a course instructor).

There were multiple methodological limitations reported in our findings, but the issue of pretests should be further explained. Although pretests are integral to many quasiexperimental designs, investigators conducting experiments with random assignment should generally avoid them (Pedhazur & Schmelkin, 1991). Unless measures are taken to control for the possible effects of the pretest (sensitization of participants to or interaction with the intervention), it is an unnecessary threat to validity.

We also had many concerns about the reporting of instruments used by the investigators to measure antigay attitudes. We were particularly skeptical about the validity of the Riddle Homophobia Scale used by Finkel et al. (2003). In general, however, the investigators of the reviewed studies did not adequately report on the validity and reliability of the instruments they used.

Overall, we applaud the investigators of the reviewed studies for their groundbreaking work in this understudied area. By drawing attention to the victimization stemming from widespread homophobia and heterosexism, perhaps funding sources will recognize the need for devoting adequate resources to continued study of this pervasive problem. The remainder of this section examines the political and ideological challenges facing researchers who pursue the development of interventions addressing antigay attitudes and concludes with applications to social work practice.

Up to this point, we have focused on the design and methodological issues involved in the intervention studies we reviewed. Our review would be incomplete, however, if we did not address the political and ideological context within which the research occurred and how this is likely to affect future investigations consistent with the reviewed studies. Specifically, it should be noted that the pervasiveness of antigay attitudes in our society, including heterosexism and homophobia, could have a chilling effect on research about these very issues. Although we are hopeful that researchers will be able to secure the funding and resources necessary to conduct well-designed and methodologically sound experiments to further this line of investigation, we suspect that there will be many obstacles.

Not everyone agrees on how homophobia and heterosexism should be addressed. A substantial and apparently politically powerful proportion of the population believes that homophobia and/or heterosexism should be embraced rather than challenged. There are numerous recent reports attesting to the ongoing culture war regarding gay and lesbian issues. For example, an Alabama lawmaker recently introduced a bill in that state’s House of Representatives that would prohibit public funds from being spent on any written material that portrays homosexuality as an acceptable lifestyle and would prohibit educators from bringing speakers or any written material to the classroom that includes content on LGB issues (Snorton, 2004).

At the national level, the political climate in the current administration seems particularly unfriendly toward the LGB population generally and specifically toward research designed to study this population. A study that examined the sexual and health risk behaviors of LGB Native Americans was one of five peer-reviewed National Institutes of Health grants targeted by Rep. Pat Toomey for defunding in 2003 (Winerman, 2004). The principal investigator of that study stated that she had been informed by political insiders that Toomey’s actions were representative of the many tactics, including the attack on gay marriage, adopted by Republicans to use the LGB population to create wedge issues in the 2004 presidential campaign (LaSala et al., 2005).

More recently, officials from the Substance Abuse and Mental Health Services Administration (SAMHSA) informed the Suicide Prevention Resource Center (SPRC) that SAMHSA’s administrator would not be attending the SPRC’s conference on suicide prevention unless conference organizers removed the words gay, lesbian, bisexual, and transgender from the title and descriptor of a planned workshop (R. Bloodworth, J. Liljeholm, & R. Vanderburgh, personal communication, February 15, 2005). The workshop presenters had to change the title from “Suicide Prevention Among Gays, Lesbians, Bisexuals, and Transgender Individuals” to “Suicide Prevention in Vulnerable Populations” in order for the workshop to be offered. Additionally, the offending words in the descriptor had to be replaced with a general reference to sexual orientation, but any reference to gender identity was not permitted.

Societal views about marginalized groups are usually slow to evolve; however, the studies reviewed herein represent attempts to speed that process. First, however,
had to progress enough to permit these studies to happen. As few as 20 years ago, studies of this kind would not be allowed on many college campuses. Indeed, not long ago, professors risked their employment at most universities if they came out (Taylor & Raeburn, 1995). From this historical perspective, the reviewed studies also represent pioneering research. We expect, however, that those espousing greater acceptance of the LGB population will continue to encounter significant resistance, particularly from the religious right.

Although official policy toward the LGB population in the field of social work is decidedly progressive and supportive, there are other indicators that suggest that these policies are not effectively put into practice. A study of gay and lesbian content in social work textbooks suggests that social work education programs continue to perpetuate heterosexist bias and discrimination (Morrow, 1996). Other studies have reported moderate to high levels of homophobia and heterosexist attitudes among social work students and little or no attention to practice with LGB populations provided in their social work training (Kriegstein, 2003; Snively et al., 2004).

There are only a few, rudimentary studies examining the effects of homophobia on social work practice. For example, homophobia among social workers and counselors has been correlated with discomfort in working with LGB clients (Hayes & Gelso, 1993; Weiner & Siegel, 1990). Missing, however, are studies examining the impact of homophobia and heterosexism on the treatment outcomes of sexual minority clients. There is also a need for research on best practices with LGB individuals and on how to prepare social workers to implement effective interventions with this population (Snively et al., 2004).

Finally, a review of the social work literature indicates a paucity of information regarding LGB issues in general (Van Voorhis & Wagner, 2002). Content analysis of the journals Social Work, Child Welfare, Social Service Review, and Families in Society for the 10 years spanning 1988 to 1997 was conducted. These particular journals were reportedly selected because of their national audience and because they are not limited to one area of social work.

Only 3.92% of the articles (77 of 1,964) addressed homosexuality at all, and the vast majority of these (more than 65%) addressed HIV or AIDS. That is, less than 1.37% of the articles addressed aspects of practice with LGB clients other than HIV or AIDS. The finding most striking to the researchers, however, was that only 4 of the 77 articles had a macro focus, and none of the non–HIV or AIDS articles focused on macro issues. The researchers concluded that the sparse coverage in these journals of issues addressing the LGB population contributed to the oppression of this group, and they questioned the commitment of the profession to its espoused principles of the ecological perspective and person-in-environment approach (Van Voorhis & Wagner, 2002).

In conclusion, it remains to be seen whether any short-term interventions can create lasting shifts in attitudes that translate into behavioral changes toward LGB individuals. It seems likely that attitude shift is a cumulative process resulting from repeated exposure to consistent information that is deemed credible. The interventions discussed herein may contribute to that process, but they should be further tested with well designed, methodologically sound experiments. This area of research should also be expanded beyond the university or classroom setting. We would like to see future research test interventions to change antigay attitudes in other settings with other populations (workplaces, community centers, churches, etc.). Most of all, however, we would like to see an overall expansion of research designed to address the pervasive victimization, discrimination, and marginalization experienced by the LGB population in this country.

REFERENCES


