Journal of Organizational Behavior

Research Article

Journal of Organizational Behavior, J. Organiz. Behav. **34**, 1076–1104 (2013)
Published online 5 November 2012 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/job.1839

A meta-analytic evaluation of diversity training outcomes

ZACHARY T. KALINOSKI, DEBRA STEELE-JOHNSON*, ELIZABETH J. PEYTON, KEITH A. LEAS, JULIE STEINKE AND NATHAN A. BOWLING

Department of Psychology, Wright State University, Dayton, Ohio, U.S.A.

Summary The purpose of this meta-analysis was to use theory and research on diversity, attitudes, and training to examine potential differential effects on affective-based, cognitive-based, and skill-based outcomes, to examine potential moderators of those effects with a focus on affective-based outcomes, and finally, to provide quantitative estimates of these posited relationships. Results from 65 studies ($N = 8465$) revealed sizable effects on affective-based, cognitive-based, and skill-based outcomes as well as interesting boundary conditions for these effects on affective-based outcomes. This study provides practical value to human resources managers and trainers wishing to implement diversity training within organizations as well as interesting theoretical advances for researchers. Practitioners have quantitative evidence that diversity training changes affective-based, cognitive-based, and skill-based trainee outcomes. This study also supports and addresses future research needs. Copyright © 2012 John Wiley & Sons, Ltd.

Keywords: diversity training; training evaluation; meta-analysis

Organizations have been citing the "business case" for diversity since the 1980s (Bendick, Egan, & Lofhjelm, 2001; Kulik & Roberson, 2008b; Page, 2007), but many researchers have noted the difficulty of obtaining the potential benefits of diversity (Edmondson & Roloff, 2009; Kulik & Roberson, 2008b; Milliken & Martins, 1996). Organizations have diversity; the challenge is to use that diversity to enhance organizational outcomes. Training can enhance organizational outcomes (e.g., Goldstein & Ford, 2002). Thus, organizations have used training as a key component of diversity initiatives, with as many as 67 percent of US organizations reporting the use of diversity training (Kulik & Roberson, 2008b). Researchers have provided evidence in qualitative reviews that diversity training has beneficial effects on learning outcomes (Kulik & Roberson, 2008a). However, there has been little empirical evidence to suggest when and under what conditions diversity training is most beneficial even though substantial information has been published advising practitioners on how to best implement diversity training (Kulik & Roberson, 2008b). Clearly, the implementation of diversity training has exceeded the pace at which researchers have been able to inform practitioners about best practices (Pendry, Driscoll, & Field, 2007).

Recently, researchers have shifted away from a main-effects model (which research results typically failed to support) to more sophisticated conceptualizations of why and when diversity training works (Kulik & Roberson, 2008b; van Knippenberg & Schippers, 2007). For example, researchers have conducted numerous narrative literature reviews (Curtis & Dreachslin, 2008; Engberg, 2004; Kulik & Roberson, 2008a; Paluck & Green, 2009; Perry, Kulik, & Field, 2009). These have focused to a greater extent on attitude change than on skill-based or knowledge change. This focus might reflect that trainers and researchers place greater importance on attitude change or a concern with understanding why diversity training has produced generally weaker effects on attitudes relative to other outcomes. Certainly, the focus in diversity training research on multiple outcomes is consistent with training

*Correspondence to: Debra Steele-Johnson, Department of Psychology, Wright State University, 3640 Colonel Glenn Highway, Dayton, Ohio 45435, U.S.A. E-mail: debra.steele-johnson@wright.edu

Received 9 May 2011
Revised 2 September 2012, Accepted 2 October 2012

research that has shown the benefits of distinguishing between training effects on affective-based, cognitive-based, and skill-based outcomes (Kraiger, Ford, & Salas, 1993). Moreover, recent research (see Bohner & Dickel, 2011, for a review) on the nature of attitudes and attitude change has introduced a discussion of the role of implicit and explicit processes. This discussion has the potential to increase our understanding of potential differential effects of diversity training on affective-based versus cognitive-based or skill-based outcomes and how one might design and deliver diversity training to strengthen effects on affective-based outcomes. Thus, we extend prior research by using theory and research on diversity, attitudes, and training to derive predicted relationships between diversity training and training outcomes. More specifically, we use a meta-analytic approach to examine potential differential diversity training effects on affective-based, cognitive-based, and skill-based outcomes and identify features incorporated in diversity training that might result in stronger effects on affective-based outcomes.

Theory and Research on Diversity Training

As noted, substantial research has focused on the effects of diversity training (e.g., Combs & Luthans, 2007; Rudman, Ashmore, & Gary, 2001; Sanchez & Medkik, 2004; Stewart, Laduke, Bracht, Sweet, & Gamarel, 2003). Briefly, some researchers have suggested that expectations surrounding diversity training should be realistic (Kulik & Roberson, 2008a). For example, diversity training might not increase minority representation within management ranks (Kalev, Dobbin, & Kelly, 2006). That is, hiring rates and staffing goals are distal outcomes of diversity training, which are achieved through a conglomeration of diversity-related practices (e.g., affirmative action plans, diversity-tolerant cultures, diversity management executives). In contrast, diversity training is likely to enhance more proximal outcomes, influencing attitudes, knowledge, and skills that might have a downstream effect on distal outcomes such as staffing (Kulik & Roberson, 2008a). Yet other research has examined outcomes of diversity training programs that vary in terms of characteristics of the training design, trainer, trainee, and training environment (Hanover & Cellar, 1998; Holladay, Anderson, Gilbert, & Turner, 2008; Holladay, Knight, Paige, & Quiñones, 2003; Holladay & Quiñones, 2005, 2008; Kath, 2004; Kulik, Pepper, Roberson, & Parker, 2007; Liberman, Block, & Uyebuko, 2010; Perry, Kulik, & Schmidtke, 1998; Robb & Doverspike, 2001; Roberson, Kulik, & Pepper, 2001). We suggest that a focus on proximal outcomes and on how we can design training to enhance those outcomes will benefit our understanding of diversity training.

Van Knippenberg and Schippers (2007) noted a distinct shift in diversity research since Williams and O'Reilly's (1998) review. That is, Williams and O'Reilly's (1998) comprehensive review of 40 years of diversity research described inconsistent research results. van Knippenberg and Schippers (2007) suggested that these inconsistent results reflected an oversimplified, main-effects approach to diversity and further suggested that the field has now shifted to a more theory-driven and better-conceptualized approach to diversity. Subsequently, there have been many attempts at furthering the theoretical basis underlying diversity training (Chavez & Weisinger, 2008; Nemetz & Christiansen, 1996; Paluck, 2006; Wiethoff, 2004).

Central to much of the theorizing about diversity and diversity training is a discussion of two perspectives: a social categorization perspective and an information-processing/decision-making perspective. Recent research has examined the implications of these perspectives, in general, associating social categorization with dysfunctional effects and the information-processing/decision-making perspective with beneficial effects of diversity (for discussion and reviews of these models, see Jackson & Joshi, 2011; Martins, Milliken, Wiesenfeld, & Salgado, 2003; van Knippenberg & Schippers, 2007). The basic premise of the social categorization perspective is that diversity could result in disruptive sub-group (i.e., in-group and out-group) formation, and the premise of the information-processing/decision-making perspective is that diversity could result in benefits from a greater array of expertise or knowledge in workgroups. van Knippenberg, De Dreu, and Homan (2004) have integrated these perspectives, suggesting that potential beneficial effects of diversity might be disrupted by social categorization if that categorization is accompanied by identity threat to produce intergroup biases. Van Knippenberg

DOI: 10.1002/job

et al. (2004) have implications for diversity training in that they suggested that researchers should focus less on the effects of a particular attribute of diversity and more on the multiple and interactive routes through which diversity attributes affect outcomes.

Integrating Training and Attitude Theory and Research with Diversity Training

Building on the theoretical foundations developed in diversity research, we seek to increase our understanding of diversity training effects by considering also theory and research on training, on attitudes, and on attitude change and prior attempts to integrate this research with diversity training. Organizations rarely conduct in-depth evaluations of their diversity training interventions (Hite & McDonald, 2006; Perry et al., 2009; Rynes & Rosen, 1995). Often, social and political pressures dictate the extent to which training is evaluated (Gutiérrez, Kruzich, Jones, & Coronado, 2000; Lynch, 1997). When organizations do evaluate, they tend to use reaction measures rather than more rigorous tools such as knowledge assessments or behavioral measurements (Bennett, 2006; Rynes & Rosen, 1995) because reaction measures are easy to develop, quick to administer, and thus cost-effective. Nonetheless, we opted to analyze diversity training from a more theoretical perspective, using the Kraiger et al. (1993) model of training evaluation. The Kraiger et al. (1993) framework identified three learning outcomes: affective-based, cognitive-based, and skill-based outcomes. Kraiger et al. (1993) defined affective-based outcomes as measures of internal states that drive perception and behavior. Affective-based outcomes include attitudes, self-efficacy, and motivation in general (Kraiger et al., 1993). Cognitive-based outcomes include verbal knowledge, knowledge organization, and cognitive strategies. Skill-based outcomes include changes in behavior.

Consistent with this multiple-outcome approach to evaluation, researchers conducting narrative literature reviews of diversity training have reported evidence of beneficial results for attitude, knowledge, and skill-based change although these results are not consistent (Curtis & Dreachslin, 2008; Engberg, 2004; Kulik & Roberson, 2008a; Paluck & Green, 2009; Perry et al., 2009). The most common outcome evaluated within diversity training is affective based (Curtis & Dreachslin, 2008), but research results are mixed. Some research has suggested that training is not a viable option for changing attitudes toward others (Bendick et al., 2001). Other researchers have found positive effects for diversity training on attitudes (Bailey, Barr, & Bunting, 2001; Hollister, Day, & Jesaitis, 1993; Pruegger & Rogers, 1994; Robb & Doverspike, 2001). However, most studies have focused on explicit attitude change, and very few studies have investigated implicit attitude change resulting from diversity training (Castillo, Brossart, Reyes, Conoley, & Phoummarath, 2007; Rudman et al., 2001).

Less research has examined effects of diversity training on cognitive-based and skill-based outcomes. However, past diversity training research has observed relatively consistent effects for change in cognitive-based outcomes, focusing primarily on the verbal knowledge attained by trainees (Gulick, Jose, Peddie, King, Kravitz, & Ferro, 2009; Holladay & Quiñones, 2008; Perry et al., 1998; Robb & Doverspike, 2001; Roberson et al., 2001; Sanchez & Medkik, 2004; Williams, 2005). Also, some research has demonstrated beneficial effects for diversity training on skill-based outcomes (Barker, 2004; Hanover & Cellar, 1998; Moyer & Nath, 1998; Perry et al., 1998; Sanchez & Medkik, 2004). This research has placed a greater focus on assessments of procedural knowledge and skill compilation than on automaticity because diversity-related behavior tends to be more complex and require more working memory capacity, relative to traditional psychomotor skills. Most of these studies have evaluated skills using self-assessments (Kulik & Roberson, 2008a).

Kulik and Roberson's (2008b) review of diversity training research provided a useful summary of this research, describing diversity training within two broad categories: training designed to disseminate information and training designed to create behavioral change. The purpose of training to disseminate information is to inform employees of the organization's diversity strategy and expectations, that is, a cognitive-based change. Training to create behavioral change can be divided further into awareness and skill training. The latter teaches trainees specific skills and behaviors, that is, a skill-based change. The former, awareness training, attempts to increase trainees' awareness

DOI: 10.1002/job

of their biases, including stereotypes. The notion is that once trainees are aware of their biases, they can change those biases, that is, an affective-based change. Kulik and Roberson (2008b) suggested that there is more substantial evidence of the success of dissemination and skill training than of awareness training in achieving training goals.

Understanding Differential Effects of Diversity Training: Implicit and Explicit Processes

We contribute to prior theory and research by suggesting that the generally stronger effects of diversity training on cognitive-based and skill-based outcomes than on affective-based outcomes might be explained in part by considering explicit and implicit processes and in part by considering access to relevant content prior to training. Research on attitude and attitude change (e.g., Gawronski & Bodenhausen, 2006) has begun to distinguish between explicit and implicit aspects of attitudes. For example, one can report one's attitude toward a target object, that is, a conscious, explicit attitude. However, one might be unable to directly report other attitudes or aspects of attitudes, that is, an implicit attitude or implicit component of an attitude. Further, research has suggested the following: (i) that measures differ in the extent to which they capture implicit versus explicit aspects of attitudes; and (ii) that explicit and implicit processes can be used to change both implicit and explicit aspects of attitudes (e.g., Briñol, Petty, & McCaslin, 2009). We draw from this research to suggest that training reflects relatively explicit processes and thus that one might expect to have more direct effects on change in explicit training outcomes. Further, we suggest that cognitive-based and skill-based outcomes reflect relatively explicit outcomes (with a few exceptions such as tacit knowledge or automaticity). However, affective-based outcomes (e.g., attitudes) to a greater extent reflect both explicit and implicit components.

Also, trainees might have greater exposure or access to content relevant to diversity attitudes than to cognitive-based or skill-based content prior to training. That is, if trainees already have favorable attitudes (explicit and implicit), diversity training would result in little attitude change. Alternatively, trainees might know prior to training which diversity attitudes are socially desirable and be able to report those attitudes explicitly. Trainees might be less likely to know prior to training specific facts or information (cognitive-based outcomes) or specific behaviors (skill-based outcomes) related to diversity. Taken together, the more explicit nature of diversity training and the potentially greater pre-training access to attitude content suggest weaker effects of diversity training on attitudes, and affective-based outcomes in general, compared with cognitive-based or skill-based training outcomes.

Hypothesis 1: Diversity training will have stronger effects on cognitive-based and skill-based outcomes, relative to affective-based outcomes.

Enhancing Diversity Training Effects on Affective-Based Outcomes: Features Related to Social Interaction and Trainee Motivation

Additionally, we contribute to prior research by examining conditions in which one might observe stronger effects on affective-based outcomes. We focus on affective-based outcomes because attitude change seems particularly important in relation to diversity training, yet diversity training has failed to reveal consistent beneficial effects on affective-based outcomes. We draw again on the notions of implicit and explicit processes and their role in attitude change (e.g., Bohner & Dickel, 2011) as well as notions that diversity attributes affect outcomes through multiple, interactive routes (i.e., social categorization and information processing/decision making; Van Knippenberg et al., 2004) to increase understanding of diversity training effects on affective-based outcomes. That is, Bohner and Dickel (2011) suggested that explicit processes affect explicit attitudes directly but also indirectly through effects

DOI: 10.1002/job

on implicit attitudes. Similarly, implicit processes affect implicit attitudes directly and through effects on explicit attitudes. In support of these contentions, research has shown that carefully processing arguments influences implicit attitudes (Briñol et al., 2009) and that individuals transform implicit associations into explicit proposition (Strack & Deutsch, 2004). (For succinctness, we refer in the following to explicit and implicit attitudes, acknowledging that there is a debate in the literature regarding whether implicit and explicit attitudes are separate attitudes or attitudes have implicit and explicit components.) We conclude, then, that responses to explicit measures are likely to be influenced in part by implicit attitudes.

These notions about how implicit and explicit processes affect attitude change are consistent with the van Knippenberg et al. (2004) model of the interactive influences of social categorization and information processing/decision making on individuals' reactions to diversity attributes and suggest features in diversity training that should enhance beneficial effects on affective-based outcomes. Van Knippenberg et al. (2004) suggested that diversity attributes have an influence on outcomes through an implicit route that is moderated by identity threat. That is, diversity attributes can cue social categorizations, that is, implicit, automatic associations, which in turn can lead to intergroup biases when identity threat is high. Further, diversity attributes have an influence on outcomes through an explicit, conscious reasoning route. These routes interact in that intergroup biases emerging from the implicit route can reduce potential beneficial effects that emerge from the explicit route.

The very nature of diversity training should cue social categorizations because the content of training is focused on race, gender, multicultural, or other aspects of diversity. Social categorizations are not inherently detrimental to outcomes; rather, individuals' reactions to the categorizations can be detrimental. Hence, the question becomes what features can be built into diversity training to improve individuals' reactions to social categorizations. Research on prejudice reduction (see Paluck & Green, 2009, for a review) has suggested that social interdependence (Deutsch, 1949) and contact under optimal conditions (Pettigrew & Tropp, 2006) can reduce prejudice. Diversity training in general likely embodies optimal conditions, including equal status, shared goals, and absence of competition. Thus, employing features that increase the opportunity to interact with others in training should benefit affective-based outcomes. These features might include using an interdependent task in training, having individuals actively participate rather than listen to lectures only, having individuals participate face to face rather than complete training on a computer, or simply providing training of greater duration. Research has demonstrated that some of these features (i.e., active vs. passive training and spaced vs. massed training) benefit cognitive-based and skill-based training outcomes (Goldstein & Ford, 2002). These features are likely to benefit affective-based outcomes as well.

Hypothesis 2: Diversity training that provides greater opportunity for social interaction will have stronger beneficial effects on affective-based outcomes, relative to training providing less opportunity for social interaction.

Whereas the preceding hypothesis addresses diversity training features that might benefit affective-based outcomes emerging through an implicit route, other features might benefit affective-based outcomes emerging through an explicit route. Specifically, van Knippenberg et al. (2004) suggested that diversity attributes have an influence on outcomes through an explicit, conscious route and that this route is moderated by task and trainee factors, including trainee motivation. Research on attitude change has described explicit processing as reflecting propositional reasoning, that is, careful consideration of attitudes and their validity (Gawronski & Bodenhausen, 2006). Moreover, Kulik and Roberson (2008b) reviewed research showing that individuals can regulate their cognitive processes relating to stereotypes and suggested that such self-regulation is likely to occur when trainee motivation is high. For example, they suggested that diversity training is likely to have stronger effects if the trainee believes the organization values diversity training. Other research has shown that training importance or relevance and trainee motivation have beneficial effects on training outcomes (e.g., Baldwin & Magjuka, 1997; Mathieu & Martineau, 1997). Thus, to the extent that trainees perceive that the stakes are high, that is, that diversity training is relevant or important, trainees should be more motivated to change their attitudes and affective-based outcomes in general.

Hypothesis 3: Diversity training will have stronger beneficial effects on affective-based outcomes, when trainee motivation is high than when it is low.

DOI: 10.1002/job

Method

Literature search

We limited our search to post-1964 (i.e., Civil Rights Act passage). We began our investigation for sources by referencing various known search engines such as PsycINFO, Business Source Complete, Social Science Citation Index, Google Scholar, SocIndex, and ERIC. In addition, we referenced narrative reviews that documented extensive literature searches (e.g., Kulik & Roberson, 2008a; Paluck & Green, 2009). We also conducted ancestry searches of identified articles to find other relevant articles for inclusion. To find relevant articles, we used the following search terms: *diversity training*, *evaluation*, *effectiveness*, *cultural diversity training*, *cultural awareness training*, *racial diversity training*, *gender diversity training*, *sexual harassment training*, *racial awareness training*, *LGBT diversity training*, *multicultural education*, and *multicultural training*.

To obtain unpublished studies, we used the aforementioned search terms for identifying unpublished documents. We located these studies by doing electronic and manual searches through past conference programs for the Society of Industrial/Organizational Psychology, Academy of Management, Association for Psychological Science, and the American Psychological Association. We also conducted electronic searches of Dissertations Abstracts and the Electronic Theses and Dissertations Center with the aforementioned search terms. We restricted the unpublished search to the past 10 years for several reasons. One, the probability of unpublished documents being published since this date was high, thus resulting in potential redundancy. Two, locating contact information for authors of older unpublished documents was less feasible. Finally, even if authors were located, the likelihood that the authors had access to the requested documents would be low. Therefore, we felt justified in limiting our search.

Inclusion criteria

The first criterion for inclusion in this meta-analysis was that the study had to provide some type of numerical data. Hence, conceptual papers and qualitative or narrative reviews discussing diversity training were excluded from further analysis. We decided to use Cohen's *d* as our effect size metric. If not provided, we used the means, standard deviations, and sample sizes from the study to calculate Cohen's *d*. We also converted inferential statistics such as univariate *F* ratios, *t* ratios, and correlations according to conversion formulas recommended by Hunter and Schmidt (2004). If the authors did not provide adequate numerical data according to the aforementioned criteria, then we discarded the study from inclusion.

All study designs (i.e., training vs. control groups with pre-tests and post-tests, training vs. control groups with post-tests only, and single training group with repeated measures) were incorporated into the meta-analysis. To combine all types of study designs to allow meaningful comparisons, we followed the advice of Morris and DeShon (2002). All three study designs use different standard deviations to calculate effect sizes. For instance, one uses the standard deviation of the change score for repeated-measures design, whereas one will use the pooled standard deviation of both the experimental and control groups in an independent-groups, post-test-only design. Cohen's *d* will be different depending on how one calculates the standard deviation. Therefore, we used transformation formulas to convert effect sizes into the same metric. Because we had more studies in the training-versus-control group designs, we elected to transform repeated measures into the independent-groups metric. To convert Cohen's *d* from repeated measures to the training-versus-control group design, we used the standard deviation of the pre-test instead of the change score. This provides an equivalent effect size to the training and control group effect sizes. Both training versus control groups with pre-tests and post-tests and training versus control groups with post-test-only designs are equivalent if the variances are homogeneous. If one cannot assume homogeneity of variance, Morris and DeShon (2002) recommended using a value that best represented the untreated population (i.e., control group *SD* and pre-test *SD*). Therefore, in cases where we could not assume homogeneity of variance, we used the pre-test standard deviation when transforming effect sizes.

DOI: 10.1002/job

From the preceding search and inclusion criteria, we collected a total of 200 articles for possible inclusion. Seventy-four of the studies were discarded because they did not contain any numerical data to analyze (e.g., conceptual articles). Sixty-one articles were discarded because they did not provide adequate data to meta-analyze. This left us with 65 usable studies with 97 total effect sizes. The total sample size was 8465. Table 1 provides a description of each study included in the meta-analysis.

Coding procedures

There were two coders used for the coding of the meta-analysis. Both coders coded all studies and are co-authors for this paper. The lead author (one of the coders) trained the other coder on coding strategies, definitions, criteria, decision rules, and other coding-related processes. This initial training lasted approximately two hours. In addition, both coders then proceeded to code an initial 10 studies. After every two to three studies, coders discussed results and clarified any conceptual discrepancies. The training of the second coder was approximately 20 hours in length, including the initial training and the coding of the initial 10 studies. Both coders then coded the remaining studies.

At the completion of the coding process, the coders met on two occasions to discuss discrepancies. The first round of discussions centered on non-systematic errors. Examples of these types of errors were miscalculations, transcription errors, missing data, and oversight. We categorized the remaining discrepancies as reflecting conceptual differences between coders. The second meeting between the coders centered on the conceptual discrepancies. The coders reached consensus on all discrepancies after discussion of the individual study in question.

Agreement indices were based on conceptual difference discrepancies (i.e., pre-second meeting data). We elected to use these data to calculate inter-rater agreement because we believed that the non-systematic error discrepancies were not based on true between-rater differences. Rather, these would be classified as random and not based on conceptual differences that each rater had about the domain of interest.

To assess inter-rater agreement, we employed two metrics. One is based on the total percentage of agreement. The average percentage of total agreement for all study variables was 89.76 percent, with type of diversity training reflecting the lowest (73.53 percent) and study setting the highest (98.53 percent) agreement. Regarding the lower level of agreement for diversity training type, about half of the cases resulted from the coders' reluctance to decide (i.e., "information not available"), whereas the other half resulted from an inability to distinguish between a single type (i.e., skill *or* awareness) and double type (i.e., skill *and* awareness). The median percentage of total agreement was 91.91 percent. The other metric that we used to assess inter-rater agreement was the ICC. We employed an ICC with a two-way random-effects model using absolute agreement. The average ICC was .854, with subject matter displaying the lowest ICC (.602) and age displaying the highest ICC (.985). Conceptual discrepancies on the subject matter variable reflected differences in understanding the "multicultural" definition. Nonetheless, both coders discussed each discrepancy individually, and both coders understood the distinctions further after discussion. The median ICC was .884. Thus, most of the total variance in variable coding was due to between-rater variance. Therefore, we felt confident that our coding procedures were reliable.

Coding of dependent variables

Affective-based outcomes

We used the Kraiger et al. (1993) model to define our outcomes. Specifically, we defined affective-based outcomes as measures of internal states that drive perception and behavior. Affective-based outcomes included measures of attitudes and motivation (including self-efficacy).

DOI: 10.1002/job

Table 1. Study descriptions.

Author(s)	Sample	Training	Design	Criteria
Badhesha, Schmidtke, Cummings, and Moore (2008)	63 undergraduate students	Religion, age	Independent groups, pre-test, post-test	Affective, cognitive
Bailey et al. (2001)	57 police officers	Disabilities	Independent groups, pre-test, post-test	Affective
Barker (2004)	62 high school students	Gender, race, age, SES	Independent groups, pre-test, post-test	Skills
Blakely, Blakely, and Moorman (1998)	175 undergraduate students	Sexual harassment	Independent groups, post-test only	Cognitive
Brathwaite (2005)[a]	76 nurses	Culture	Repeated measures	Cognitive
Brown (2004, Cohort 1)	60 undergraduate students	Race, ethnicity, SES, culture, gender, religion	Repeated measures	Affective
Brown (2004, Cohort 2)	49 undergraduate students	Race, ethnicity, SES, culture, gender, religion	Repeated measures	Affective
Bush and Ingram (2001)	91 sales and marketing professionals	Generic	Repeated measures	Skills
Byrnes and Kiger (1990)	164 undergraduate students	Race	Independent groups, pre-test, post-test	Affective
Castillo et al. (2007)	84 counseling graduate students	Multicultural	Independent groups, pre-test, post-test	Affective, cognitive, skills
Celik, Abma, Klinge, and Widdershoven (2011)	31 nurses	Generic	Repeated measures	Cognitive
Chiasson (2006)	353 undergraduate students	LGBT	Independent groups, pre-test, post-test	Affective
Combs and Luthans (2007)	260 mixed employees (government, insurance, manufacturing)	Generic	Independent groups, post-test only	Affective, skills
Cornett-DeVito and McGlone (2000)	40 police officers	Generic	Repeated measures	Skills
Davis (2008)	49 government employees (engineers, systems analysts, facilities managers)	Generic	Independent groups, post-test only	Cognitive
De Meuse, Hostager, and O'Neill (2007)	35 manufacturing managers	Generic	Repeated measures	Affective
Dean (2008)	210 undergraduate students	Race and gender	Independent groups, post-test only	Affective
Dickson, Argus-Calvo, and Tafoya (2010)	60 counseling graduate students	Race	Independent groups, pre-test, post-test	Affective, cognitive, skills
Doorenbos and Schim (2004)	91 hospice employees	Information unavailable	Independent groups, post-test only	Skills

(*Continues*)

DOI: 10.1002/job

Table 1. (Continued)

Author(s)	Sample	Training	Design	Criteria
Edelstein (2007)	178 nutritionists	Generic	Repeated measures	Affective
Flavin (1997)	11 hospice nurses	Race, LGBT	Repeated measures	Cognitive
Gany and de Bocanegra(1996)	80 maternity infant care center employees (clerical, nurses, doctors)	Multicultural	Repeated measures	Affective, cognitive
Gulick et al. (2009)	200 undergraduate students	Generic	Independent groups, post-test only	Cognitive, skills
Guy-Walls (2007)	150 undergraduate students	Information unavailable	Independent groups, post-test only	Cognitive
Hanover and Cellar (1998)	61 research and engineering managers	Generic	Independent groups, pre-test, post-test	Affective, skills
Hauenstein, Findlay, and McDonald (2010, Study 1)	46 military equal-opportunity advisors	Information unavailable	Repeated measures	Skills
Hauenstein et al. (2010, Study 2)	55 military equal-opportunity advisors	Information unavailable	Repeated measures	Skills
Hayes et al. (2004)	90 substance abuse counselors	Age, race, religion, LGBT, ethnicity, gender	Independent groups, pre-test, post-test	Affective
Hill and Augoustinos (2001)	62 judiciary and court employees	Ethnicity	Repeated measures	Affective, cognitive
Holladay et al. (2008)	165 healthcare managers	Generic	Repeated measures	Affective, cognitive, skills
Hostager and De Meuse (2008)	302 undergraduate students	Age, disabilities, race, gender	Independent groups, pre-test, post-test	Affective
Jefferson (2001)	261 undergraduate students	Generic	Repeated measures	Affective
Krings, Bollmann, and Palazzo (2009)	87 MBA students	Gender, race, LGBT, nationality	Repeated measures	Affective
Kulik et al. (2007, Study 2)	110 graduate research assistants	Generic	Independent groups, post-test only	Skills
Landis, Brislin, and Hulgus (1985)	45 undergraduate students	Race	Independent groups, post-test only	Cognitive
Luger (2011)	24 undergraduate students	Generic	Repeated measures	Affective, cognitive, skills
Majumdar, Keystone, and Cuttress (1999)	48 medical residents	Multicultural	Independent groups, pre-test, post-test	Affective, cognitive, skills
Mak and Buckingham (2007)	142 undergraduate students	Ethnicity	Independent groups, pre-test, post-test	Affective, skills

(*Continues*)

DOI: 10.1002/job

Table 1. (Continued)

Author(s)	Sample	Training	Design	Criteria
Mausehund, Timm, and King (1995)	67 undergraduates students	Generic	Independent groups, pre-test, post-test	Cognitive
Moyer and Nath (1998, Study 1)	60 undergraduate students	Sexual harassment	Independent groups, post-test only	Skills
Moyer and Nath (1998, Study 2)	84 undergraduate students	Sexual harassment	Independent groups, post-test only	Skills
Neville and Furlong (1994)	73 undergraduate students	Race	Independent groups, post-test only	Affective, skills
Parker, Moore, and Neimeyer (1998)	54 counseling graduate students	Race and ethnicity	Independent groups, pre-test, post-test	Affective
Pedersen (2010)	45 undergraduate students	Cross-cultural	Independent groups, pre-test, post-test	Affective
Perry et al. (1998)	34 undergraduate students	Sexual harassment	Independent groups, post-test only	Cognitive, skills
Preusser, Bartels, and Nordstrom (2011, Cohort 1)	35 university employees	Sexual harassment	Repeated measures	Affective, cognitive, skills
Preusser et al. (2011, Cohort 2)	35 university employees	Sexual harassment	Repeated measures	Affective, cognitive, skills
Probst (2003)	94 undergraduate students	Multicultural	Independent groups, pre-test, post-test	Affective
Pruegger and Rogers (1994)	64 undergraduate students	Generic	Independent groups, pre-test, post-test	Affective
Randolph, Landis, and Tzeng (1977)	35 undergraduate students	Race	Independent groups, post-test only	Affective, cognitive, skills
Reynolds (2010)	18 taxi drivers	Age, disabilities	Repeated measures	Cognitive
Robb and Doverspike (2001)	90 undergraduate students	Sexual harassment	Independent groups, post-test only	Affective
Robinson and Bradley (1997)	44 undergraduate students	Multicultural	Independent groups, pre-test, post-test	Affective, cognitive
Rudman et al. (2001, Study 1)	35 undergraduate students	Race	Independent groups, pre-test, post-test	Affective
Rudman et al. (2001, Study 2)	86 undergraduate students	Race	Independent groups, pre-test, post-test	Affective

(*Continues*)

DOI: 10.1002/job

Table 1. (Continued)

Author(s)	Sample	Training	Design	Criteria
Sanchez and Medkik (2004)	125 county government supervisors and managers	Generic	Independent groups, post-test only	Cognitive, skills
Scher (2008)	37 clinical psychology doctoral students	LGBT	Repeated measures	Affective, cognitive
Soble, Spanierman, and Liao (2011)	138 undergraduate students	Race	Independent groups, pre-test, post-test	Affective
Steinfeldt and Wong (2010)	43 counseling graduate students	Race	Independent groups, pre-test, post-test	Affective
Stewart et al. (2003)	24 undergraduate students	Race	Independent groups, post-test only	Affective
Tang, Hernandez, and Adams (2004)	12 medical students	Generic	Repeated measures	Affective
Thomas (2008)	89 IT professionals	Information unavailable	Independent groups, pre-test, post-test	Affective, skills
Thomas and Cohn (2006)	47 healthcare professionals	Race, culture, religion	Repeated measures	Affective
Williams (2005)	47 social workers	Race, culture, ethnicity	Independent groups, pre-test, post-test	Affective, cognitive, skills
Zarubin (2008)	242 graduates	Generic	Repeated measures	Affective, skills

Note: SES, socioeconomic status.
[a]Study that has been identified as an outlier (i.e., $d = 3.11$).

Cognitive-based outcomes
Relying on the Kraiger et al. (1993) model, we defined cognitive-based outcomes as measures of verbal knowledge, knowledge organization, and cognitive strategies.

Skill-based outcomes
Again relying on the Kraiger et al. (1993) model, we defined skill-based outcomes as measures of behavior and behavioral intentions.

Coding of moderators: features related to social interaction

We identified the following training features as reflecting to some extent opportunities for social interaction. We employed multiple decision rules for coding moderators using the sub-group method. Also, we coded continuous moderator variables (e.g., total number of training hours) using both the sub-group method based on theory and/or the distribution of values as well as treating them as continuous variables and analyzing them with bivariate correlations.

Task interdependence
Task interdependence was coded as "yes" if the nature of the training reflected interactions with others or an interdependent task or goals and "no" if the nature of training did not include these features.

DOI: 10.1002/job

Forms of instruction
Each study had a relatively unique combination of forms of instruction, so we coded each study as reflecting either "active," "passive," or "both" components. Examples of active learning methods are role plays, simulations, discussions, games, etc. Examples of passive learning methods are lecture, reading, video, pictures, demonstrations, etc. We coded this variable as "both" if both methods were used.

Training medium
The training medium was coded as "instructor" if a person delivered the training and "computer" if the training was mediated through a computer without the assistance of a human instructor (e.g., an online training program).

Training duration
Training duration was coded as less than 4, 4–8, 12–16, and more than 16 hours. No studies reported a duration of 8–12 hours; hence, there is no sub-group for this interval. We formed these four, discrete sub-groups on the basis of their distribution across all meta-analytic studies. We also analyzed training duration as a continuous variable, that is, as number of hours. Correlations between duration coded as a continuous variable and diversity outcomes were not significant and are not addressed further in our results.

Practice spacing
Practice spacing was coded as "massed" if the training was completed in one session and as "distributed" if the training was spread out over multiple sessions.

Coding of moderators: features related to trainee motivation

We identified the following features as reflecting to some extent features that relate to trainee motivation, in particular considering characteristics of the trainer, trainee, and setting that might influence trainees' perceptions that diversity training was important and relevant. We used the same coding rules and procedures as described earlier.

Trainer source
Trainer source reflected the nature and level of employment of the individual(s) who delivered the training. An "internal manager" trainer was someone who managed or supervised trainees, in either a work or educational context. "Internal other" reflected someone in the organization other than a manager/supervisor (e.g., diversity/inclusion officer, human resources (HR) representative, and learning and development coordinator). An "external" trainer was someone outside the organization hired on a contractual basis to deliver diversity training.

Type of participant
We coded type of participant as employee or student.

Training setting
We coded setting as lab or field. Type of participant was moderately, but not strongly, correlated with setting ($r = .41$), so we examined each variable separately.

Coding of control variables

We coded a number of other potential moderators of diversity training effects as a way of more completely describing the nature of the studies included in this meta-analysis and examining whether diversity training results were influenced by these factors.

DOI: 10.1002/job

Needs assessment
We coded the needs assessment variable as "yes" or "no," depending on whether the study contained details about conducting a needs assessment or explicitly mentioned the lack of a needs assessment.

Training or education
We coded King, Gulick, and Avery's (2010) distinction between diversity training and diversity education. They distinguished between the two on the basis of setting (i.e., organizational vs. educational) and content (i.e., needs assessment, context, competency development, and demonstrations/practice vs. frequent/structured feedback, performance metrics, and exposure to affective/cognitive processes).

Training choice
Training choice was coded as voluntary if the study reported that participation was optional or under individual discretion. We coded the study as mandatory if participation *in the study* was voluntary but completing the diversity training was mandatory.

Training type
We coded training type as awareness only or awareness plus skill training. Using Kulik and Roberson's (2008b) terminology, we coded a study as "awareness" if the training expanded beyond dissemination of information but only to focus on creating awareness of and/or altering cognitive biases. We coded a study as "skill/awareness" if the training focused on both awareness and skills training (i.e., changing behavior or intentions).

Subject matter
We coded subject matter into five distinct sub-groups. Any study that focused on one demographic characteristic only (e.g., race, LGBT, disabilities) was coded as "single," acknowledging that the most frequent single characteristic was race. If authors described a generic or broad-based approach to diversity, we coded it as "generic." If a study described the training content as learning about multiple cultures, we coded it as "multicultural." If a study discussed multiple, specific forms of diversity (i.e., race and gender), we coded it as "combination." Finally, if a study focused on "sexual harassment," we coded this study as such.

Compliance/legal content
We coded this as "yes" if there was any mention of compliance or legal components in the training content and "no" if these factors were not mentioned.

Training focus
We coded training focus as "differences" if the primary focus of the training was on emphasizing the differences between individuals and as "similarities/differences" if both components were mentioned. Almost no studies focused only on "similarities."

Types of diversity
We coded the percent of female trainees and the percent of Caucasian trainees variables into three sub-groups (i.e., less than 40 percent women/Caucasian, between 40 and 60 percent women/Caucasian, and greater than 60 percent women/Caucasian). The 40–60 percent interval represented greater diversity (e.g., approximately equal numbers of men and women), whereas the greater than 60 percent and less than 40 percent intervals represented less diversity. The specific cut-off points reflected the distribution of values observed across all meta-analytic studies rather than any pre-determined conceptual rule. We also analyzed these two variables as continuous moderator variables.

We coded trainer race as "Caucasian," "non-Caucasian," or "both" (i.e., multiple trainers of different racial backgrounds) and trainer gender as male or female.

DOI: 10.1002/job

Self-report
If a study asked participants to report or judge their own skill level or competence, we coded it as self-report. If participants completed a test (e.g., of topic knowledge) or were observed, we coded that study as non-self-report.

Time of evaluation
We coded the time of evaluation as the time (in months) between the end of training and when training was evaluated. We created three sub-groups for this variable (i.e., end of training, less than or equal to one month post-training, and greater than one month post-training). Most (76 percent) studies used end-of-training evaluations. Also, because one might expect differences between a week and a year time lags, we elected to divide the post-training evaluations into two separate sub-groups: any time lag less than or equal to one month and any time lag more than one month.

Study rigor
We coded a study that included pre-tests and post-tests for one group only, that is, a group receiving diversity training, as training group only repeated measures. We coded a study that compared a group receiving diversity training with a control group on a post-test but included no pre-test as training versus control post-test only. Finally, we coded a study that compared a group receiving diversity training with a control group and included both pre-tests and post-tests as training versus control pre-test and post-test.

Random assignment
We coded whether the study used a quasi-experimental versus experimental design.

Publication status and year
Finally, we coded publication status as published versus unpublished. We analyzed publication year as a continuous variable only. We were unable to code this variable using the sub-group method because most studies were published in the last 20 years (14 studies in the 1990s and 50 from 2000 to the present) and almost no studies in earlier decades (zero studies in the 1960s, one study in the 1970s, one study in the 1980s).

Independence of observations

Following the recommendations of Hunter and Schmidt (2004), we accounted for non-independence in our data by averaging multiple indices of the same constructs originating from a single study. If there were multiple data points reflecting different constructs (e.g., cognitive-based vs. skill-based outcomes), then these were deemed to be independent despite their origin from the same sample. Considering that we were interested in the different outcomes separately, we felt justified in not aggregating these variables further. This procedure was also consistent with recent training meta-analyses (e.g., Arthur, Bennett, Edens, & Bell, 2003).

Outliers

Following the procedures outlined by Huffcutt and Arthur (1995) and Arthur et al. (2003), we sought to detect outliers in the data set by computing deviancy statistics of the individual Cohen's *d*. Deviancy statistics allowed us to determine the extent to which each Cohen's *d* deviated from the average, meta-analytically derived Cohen's *d*. Once we computed these statistics, we took the absolute value of the deviancy statistics (i.e., sample-adjusted meta-analytic deviancy), rank-ordered them, and subjected them to a scree plot analysis. Just as one would determine the number of factors in a factor analysis, we determined the number of outliers by the point at which the scree plot leveled off after a steep decline. Our cut-off (sample-adjusted meta-analytic deviancy ≈ 10.20) resulted in the detection of only one outlier (i.e., $d = 3.11$). Thus, we were left with 96 analyzable effect sizes for this meta-analysis. However, we analyzed the

DOI: 10.1002/job

results with and without the outlier, so readers can discern the differences between the results with or without the outlier. The outlier study was conducted by Brathwaite (2005) using a sample of registered nurses. The only discernible reason for such a large effect size compared with other studies is the small pre-test standard deviation ($SD = 0.18$). We did not observe a similarly sized standard deviation in any of the other studies included in this meta-analysis. Thus, a modest mean difference on a 4-point scale (i.e., $3.38 - 2.82 = 0.56$) resulted in a very large strength of effect given the small observed standard deviation. Generally, the inclusion of this outlier skewed the average and corrected effect size and standard deviation for cognitive outcomes upward ($\Delta\delta = 0.12$, $\Delta SD = 0.25$). This change becomes even larger when k decreases. Thus, given the drastic changes in average effect size with the inclusion of this study, we recommend readers to interpret the results without the outlier.

Analyses

We followed the meta-analysis approach recommended by Hunter and Schmidt (2004). We corrected for sampling error by weighting the average Cohen's d by its sample size. We then corrected for criterion unreliability to estimate a "true" population parameter. For every effect size, we reported an internal consistency alpha. For studies that failed to report alphas for their criterion measure, we created artifact distributions. That is, we computed an average alpha from the other reported alphas for the given outcome. Finally, we computed confidence intervals (CI) and credibility intervals around the average corrected effect sizes to ascertain the level of significance and moderation, respectively. We used a random-effects model for both main-effect and moderation analyses.

We tested moderators using the sub-grouping method for all moderator variables. In addition, we followed the recommendations of Hunter and Schmidt (2004) regarding the testing of continuous moderator variables. Hunter and Schmidt (2004) recommended using bivariate correlations between moderator variables and effect sizes of each outcome *only if* the variables are not highly intercorrelated. Therefore, we tested the intercorrelations of all five continuous moderator variables. Only two of the 10 possible correlations were significantly related, and these correlations were modest. Specifically, the percentage of Caucasians was negatively related to the year of publication ($r = -.39$) and the percentage of women was positively related to the number of training hours ($r = .33$). Therefore, we elected to use bivariate correlations to test the effects of these moderators directly on the effect sizes of our outcomes.[1]

Results

We calculated an overall, corrected effect size for diversity training across all criterion domains ($k = 96$, $N = 8389$, $\delta = 0.43$, $SD_\delta = 0.33$, 95 percent CI = 0.35–0.43). Thus, overall, diversity training had between small and medium effects on learning (Table 2). One should interpret this estimate with caution because it violates the non-independence assumption by including multiple data points from the same samples. Therefore, we analyzed and interpreted effects and moderators of diversity training on the three training outcomes. To detect moderators, we interpreted the 80 percent credibility intervals and the percentage of variance accounted for by artifacts. Wide credibility intervals as well as less

[1]Steel and Kammeyer-Mueller (2002) recommended using weighted least squares regression to test for continuous variable moderation in meta-analysis. We decided against this approach because weighted estimates become unstable when the primary study sample size is small. Given our average and median sample size (mean = 89.56, median = 62.00), we felt that using bivariate correlations was our best option. To exemplify this point, Steel and Kammeyer-Mueller (2002) suggested that when $r = .99$ with $N = 20$ and $r = .60$ with $N = 20\,000$, both studies would be given equal weights. In addition, Hunter and Schmidt (2004) pointed out that Steel and Kammeyer-Mueller's (2002) Monte Carlo study focused on the impact of four separate tests on multiple R, not on the beta weights.

DOI: 10.1002/job

Table 2. Diversity training effects on affective-based, cognitive-based, and skill-based outcomes.

Variable	*k*	*N*	*d*	SD_d	δ	SD_δ	95% CI	80% CV	%VAR
Overall	96	8389	0.39	0.37	0.43	0.33	[0.35, 0.43]	[0.01, 0.85]	34.90
Overall (with outlier)	97	8465	0.41	0.44	0.46	0.43	[0.37, 0.45]	[−0.09, 1.01]	23.88
Affective based	44	4267	0.27	0.30	0.30	0.23	[0.21, 0.33]	[0.01, 0.59]	49.21
Attitudes	40	3653	0.23	0.28	0.26	0.20	[0.16, 0.30]	[0.00, 0.51]	58.12
Self-efficacy	4	614	0.53	0.26	0.55	0.21	[0.37, 0.69]	[0.29, 0.82]	40.77
Cognitive based	25	1724	0.62	0.48	0.71	0.46	[0.52, 0.72]	[0.11, 1.30]	27.06
Cognitive based (with outlier)	26	1800	0.72	0.63	0.83	0.71	[0.62, 0.82]	[−0.09, 1.74]	13.63
Knowledge (with outlier)	23	1671	0.73	0.71	0.83	0.75	[0.63, 0.83]	[−0.12, 1.79]	12.08
Knowledge	22	1595	0.62	0.50	0.70	0.49	[0.52, 0.72]	[0.07, 1.33]	23.82
Structure	3	129	0.64	0.03	0.64	0.00	[0.28, 1.00]	[0.64, 0.64]	100.00
Skill based	27	2398	0.43	0.29	0.47	0.24	[0.35, 0.51]	[0.16, 0.78]	49.23
Skill tests	19	1390	0.54	0.27	0.60	0.22	[0.43, 0.65]	[0.31, 0.89]	60.86
Behavioral intentions	2	295	0.15	0.32	0.18	0.31	[−0.08, 0.38]	[−0.22, 0.57]	27.16
On-the-job behaviors	6	713	0.35	0.19	0.37	0.07	[0.20, 0.50]	[0.28, 0.45]	89.18

Note: *k* = number of studies; *N* = total sample size across all effect sizes; *d* = the sample size weighted mean observed effect size; SD_d = the sample size weighted standard deviation; δ = sample size weighted mean observed effect size corrected for unreliability; SD_δ = standard deviation of average corrected effect size; 95% CI = 95 percent confidence interval; 80% CV = 80 percent credibility interval; %VAR = the percent of observed variance attributable to all artifacts; (with outlier) = analysis run with the study containing the outlier included.

than 75 percent of variance in effect sizes accounted for by artifacts were our criteria for determining the presence of moderators. We used the criteria of zero to minimally overlapping (0.01–0.02) CIs to determine significant differences between effect sizes.

Diversity training effects on affective-based, cognitive-based, and skill-based outcomes (Hypothesis 1)

Diversity training exhibited a positive effect on the affective-based, cognitive-based, and skill-based outcomes, and the 95 percent CIs did not include zero (Table 2). Diversity training exhibited a medium to large effect on cognitive-based and skill-based outcomes, whereas it exhibited only a small-sized to medium-sized effect on affective-based outcomes. Also, the CIs for cognitive-based (δ = 0.71, 95 percent CI = 0.52–0.72) and skill-based (δ = 0.47, 95 percent CI = 0.35–0.51) outcomes did not overlap with affective-based (δ = 0.30, 95 percent CI = 0.21–0.33) outcomes, and the effect sizes were larger. The effect sizes for cognitive-based and skill-based outcomes did not differ significantly. Therefore, we found support for Hypothesis 1.

In addition, to more completely describe our data, we examined finer-grained distinctions. Specifically, we looked at the differences between attitudes and self-efficacy for affective-based outcomes and found that diversity training had a significantly larger effect on trainee self-efficacy (δ = 0.55) than on trainee attitudes (δ = 0.26). Also, tests of diversity-related skills (δ = 0.60) exhibited a larger effect size compared with measures of behavioral intention (δ = 0.18). This finding may be due to low sample size on the behavioral intention measures. All other category distinctions did not differ significantly from each other (Table 2).

Finally, we examined two statistical tests (i.e., 80 percent credibility interval and percentage of variance accounted for by artifacts) to determine whether moderators existed within our outcomes. For all three outcomes, the credibility intervals were very wide (at least 0.51 for affective-based outcomes), and the variance accounted for by artifacts was less than 50 percent for all three outcomes. Therefore, we felt confident in pursuing our tests of moderators.

DOI: 10.1002/job

Effects of moderators: features related to social interaction (Hypothesis 2)

We predicted that diversity training that provided greater opportunity for social interaction would have more beneficial effects on affective-based outcomes. We examined five features of training relating to opportunity for social interaction: task interdependence, forms of instruction, training medium, training duration, and practice spacing. (Although not predicted, for completeness, we also reported significant effects for these features as moderators of training effects on cognitive-based outcomes and noted that there were no significant moderator effects on skill-based outcomes.)

Results provided support for Hypothesis 2 (Table 3). Diversity training including features that reflected greater opportunity for social interaction yielded stronger effects on affective-based outcomes. Specifically, diversity training incorporating task interdependence ($\delta = 0.41$) had a significantly larger effect on affective-based outcomes than when no task interdependence was incorporated ($\delta = 0.14$). Training using both active and passive forms of instruction ($\delta = 0.37$) yielded a larger effect size, compared with those using passive instruction only ($\delta = 0.07$; CI included zero). If training was mediated by a computer ($\delta = 0.08$), the effect size was much smaller than if the training was in person, mediated by a human instructor ($\delta = 0.32$). Also, the CI for computer-mediated training included zero. Training duration that was four hours or more yielded larger effect sizes (δs = 0.52, 0.49, and 0.46) than training that was less than four hours ($\delta = 0.11$) for affective-based outcomes. More importantly, the CI for training that was less than four hours

Table 3. Diversity training features of related to social interaction (Hypothesis 2).

Variable	*k*	*N*	*d*	SD_d	δ	SD_δ	95% CI	80% CV	%VAR
Affective-based outcomes									
Task interdependence									
Yes	27	2173	0.38	0.31	0.41	0.22	[0.30, 0.46]	[0.13, 0.69]	55.25
No	12	1782	0.12	0.23	0.14	0.17	[0.03, 0.21]	[−0.07, 0.36]	54.42
Forms of instruction									
Active	9	889	0.27	0.35	0.32	0.30	[0.14, 0.40]	[−0.07, 0.71]	36.02
Passive	8	833	0.07	0.22	0.07	0.10	[−0.07, 0.21]	[−0.06, 0.21]	81.97
Both	24	2384	0.34	0.27	0.37	0.19	[0.26, 0.42]	[0.12, 0.61]	56.32
Training medium									
Instructor	40	3941	0.29	0.29	0.32	0.22	[0.28, 0.35]	[0.04, 0.60]	50.14
Computer	4	326	0.07	0.28	0.08	0.20	[−0.15, 0.29]	[−0.18, 0.34]	61.61
Training duration									
Less than four hours	13	1553	0.09	0.19	0.11	0.02	[−0.01, 0.19]	[0.09, 0.13]	99.41
Four to eight hours	5	576	0.49	0.22	0.52	0.10	[0.32, 0.66]	[0.39, 0.64]	80.19
12–16 hours	5	574	0.45	0.19	0.49	0.05	[0.28, 0.62]	[0.42, 0.55]	94.17
16+ hours	7	440	0.42	0.40	0.46	0.33	[0.23, 0.61]	[0.04, 0.89]	41.59
Practice spacing									
Massed	19	2358	0.19	0.27	0.21	0.21	[0.11, 0.27]	[−0.05, 0.47]	47.89
Distributed	20	1484	0.41	0.33	0.45	0.24	[0.31, 0.51]	[0.14, 0.76]	52.82
Cognitive-based outcomes									
Task interdependence									
Yes	17	934	0.80	0.47	0.91	0.42	[0.67, 0.93]	[0.37, 1.44]	36.62
No	6	591	0.28	0.39	0.33	0.41	[0.12, 0.44]	[−0.20, 0.86]	26.04
Forms of instruction									
Active	5	349	0.75	0.11	0.82	0.00	[0.53, 0.97]	[0.82, 0.82]	100.00
Passive	6	451	0.35	0.49	0.41	0.57	[0.16, 0.54]	[−0.31, 1.14]	21.70
Practice spacing									
Massed	10	944	0.46	0.34	0.54	0.33	[0.33, 0.59]	[0.13, 0.96]	36.93
Distributed	13	707	0.84	0.57	0.92	0.52	[0.69, 0.99]	[0.25, 1.59]	26.17

Note: *k* = number of studies; *N* = total sample size across all effect sizes; *d* = the sample size weighted mean observed effect size; SD_d = the sample size weighted standard deviation; δ = sample size weighted mean observed effect size corrected for unreliability; SD_δ = standard deviation of average corrected effect size; 95% CI = 95 percent confidence interval; 80% CV = 80 percent credibility interval; %VAR = the percent of observed variance attributable to all artifacts.

DOI: 10.1002/job

included zero. The correlation between training duration and affective-based outcomes was modest ($r_{corrected} = .30$) and non-significant, but this could be due to low power (i.e., $k = 31$). Finally, distributed practice ($\delta = 0.45$) exhibited stronger effects on affective-based outcomes compared with massed practice ($\delta = 0.21$).

Although not addressed in our predictions, we also observed that diversity training effects on cognitive-based outcomes were moderated by task interdependence ($\delta = 0.91$ vs. 0.33, for interdependent vs. not), forms of instruction (active $\delta = 0.82$ vs. passive $\delta = 0.41$), and practice spacing (distributed $\delta = 0.92$ vs. massed $\delta = 0.54$).

Effects of moderators: features related to trainee motivation (Hypothesis 3)

We predicted that diversity training that reflected to some extent features that relate to trainee motivation would have more beneficial effects on affective-based outcomes. In particular, we examined characteristics of the trainer, trainee, and setting that might influence trainees' perceptions that diversity training was important and relevant, that is, that might influence trainee motivation. We examined trainer source, type of participant, and training setting. (Although not predicted, for completeness, we also reported significant effects for these features as moderators of training effects on cognitive-based outcomes and noted that there were no significant moderator effects on skill-based outcomes.)

Results provided support for Hypothesis 3 (Table 4). Diversity training including features that related to higher trainee motivation had stronger effects on affective-based outcomes. Specifically, the effect sizes for affective-based outcomes were larger when the trainer was a direct manager/supervisor ($\delta = 0.44$) than other internal staff members (e.g., diversity/inclusion manager, HR generalist, $\delta = 0.05$). Also, the effect size for other internal staff included zero. Similarly, employee participants tended to benefit to a greater extent from diversity training ($\delta = 0.50$) than student

Table 4. Study features related to trainee motivation (Hypothesis 3).

Variable	*k*	*N*	*d*	SD_d	δ	SD_δ	95% CI	80% CV	%VAR
Affective-based outcomes									
Trainer source									
Internal manager	13	1038	0.40	0.24	0.44	0.06	[0.28, 0.52]	[0.36, 0.52]	94.22
Internal other	6	710	0.04	0.16	0.05	0.00	[−0.11, 0.19]	[0.05, 0.05]	100.00
External	6	809	0.30	0.30	0.32	0.26	[0.16, 0.44]	[−0.01, 0.65]	33.63
Both	2	151	0.31	0.05	0.34	0.00	[−0.01, 0.63]	[0.34, 0.34]	100.00
Type of participant									
Student	29	2978	0.19	0.27	0.22	0.20	[0.12, 0.26]	[−0.04, 0.47]	53.87
Employee	15	1289	0.46	0.26	0.50	0.15	[0.35, 0.57]	[0.31, 0.68]	72.78
Training setting									
Lab	7	627	0.06	0.33	0.07	0.26	[−0.10, 0.22]	[−0.27, 0.41]	42.67
Field	36	3567	0.31	0.28	0.35	0.20	[0.24, 0.38]	[0.10, 0.60]	56.27
Cognitive-based outcomes									
Type of participant									
Student	13	1018	0.50	0.49	0.57	0.47	[0.38, 0.62]	[−0.03, 1.17]	23.30
Employee	12	706	0.79	0.41	0.90	0.38	[0.64, 0.94]	[0.41, 1.39]	40.40
Trainer source									
Internal manager	4	363	0.45	0.66	0.54	0.69	[0.24, 0.66]	[−0.35, 1.43]	11.03
Internal other	3	134	1.31	0.25	1.61	0.00	[0.93, 1.69]	[1.61, 1.61]	100.00
Study setting									
Lab	6	552	0.39	0.28	0.46	0.22	[0.22, 0.56]	[0.18, 0.73]	57.01
Field	19	1172	0.73	0.51	0.82	0.49	[0.61, 0.85]	[0.19, 1.45]	26.68

Note: k = number of studies; N = total sample size across all effect sizes; d = the sample size weighted mean observed effect size; SD_d = the sample size weighted standard deviation; δ = sample size weighted mean observed effect size corrected for unreliability; SD_δ = standard deviation of average corrected effect size; 95% CI = 95 percent confidence interval; 80% CV = 80 percent credibility interval; %VAR = the percent of observed variance attributable to all artifacts.

DOI: 10.1002/job

participants ($\delta = 0.22$). Finally, field studies ($\delta = 0.35$) showed stronger effects than lab studies ($\delta = 0.07$) for affective-based outcomes, and the CI for lab studies included zero.

Although not addressed in our predictions, we mention effects for cognitive-based outcomes as well. That is, we observed a stronger effect for internal other trainers ($\delta = 1.61$) than internal manager trainers ($\delta = 0.54$), opposite the effect observed for affective-based outcomes. However, similar to affective-based outcomes, we observed a stronger effect for employee participants ($\delta = 0.90$) than for student participants ($\delta = 0.57$) and a stronger effect for field ($\delta = 0.82$) than for lab settings ($\delta = 0.46$).

Effects of control variables in diversity training outcomes

Finally, we examined the role of a number of control variables to more completely understand diversity training effects. We reported significant effects for any control variable that moderated diversity training effects (Table 5). As stated earlier, for completeness, we reported effects on cognitive-based outcomes and the one observed effect on skill-based outcomes.

We first examined aspects of training design and content. There were not enough data to analyze the presence versus absence of a needs assessment for affective-based outcomes only, but we did not observe a significant effect. Similarly, we observed no significant differences in effect sizes for training choice (voluntary vs. mandatory), training type (awareness and skill vs. awareness only), compliance/legal content, or training focus (differences only vs. similarities and differences). However, we did observe that diversity education ($\delta = 0.45$) had a stronger effect on affective-based outcomes than did diversity training ($\delta = 0.26$). Also, diversity training with subject matter focused on a single attribute of demographic diversity (e.g., race) yielded larger effects ($\delta = 1.28$) on cognitive-based outcomes than generic ($\delta = 0.52$), multicultural ($\delta = 0.54$), or sexual harassment ($\delta = 0.54$) forms of diversity training.

Next, we examined aspects of participants and trainers, that is, participant and trainer race and gender, to assess their role in diversity training effects. We observed larger training effects on cognitive-based outcomes for training groups that were less than 40 percent Caucasian ($\delta = 1.28$), compared with training groups that were greater than 60 percent Caucasian ($\delta = 0.42$). Similarly, we observed larger training effects on cognitive-based outcomes for training groups that were greater than 60 percent women ($\delta = 0.71$), compared with training groups that were less than 40 percent women ($\delta = 0.38$). Further, we observed a larger training effect on cognitive-based outcomes for Caucasian trainers ($\delta = 1.61$) than for non-Caucasian trainers ($\delta = 0.54$), but this was based on just four studies. We observed no effect for trainer gender.

Finally, we examined aspects of measures used in training, evaluation of training, and quality of the research study. Not surprisingly, we observed significantly larger effect sizes for self-report measures ($\delta = 0.85$) than for non-self-report measures ($\delta = 0.60$) on cognitive-based outcomes. We did not analyze this variable for affective-based outcomes because all affective-based measures were self-report. Similarly, we observed that studies with end-of-training evaluation ($\delta = 0.82$) exhibited significantly larger effects for cognitive-based outcomes than studies with time lags greater than one month ($\delta = 0.39$). Also, in relation to study rigor, studies using training group only with repeated-measures designs ($\delta = 1.18$) yielded larger effect sizes than training versus control post-test only ($\delta = 0.45$) and training versus control pre-test and post-test measures ($\delta = 0.66$) on cognitive-based outcomes. Moreover, we observed stronger effects for published studies ($\delta = 0.38$) than unpublished studies ($\delta = 0.13$) on affective-based outcomes. Year of publication was significantly correlated with effect size only for skill-based outcomes, but the implication is that the size of the impact of diversity training increases every year.[2] Finally, random assignment was unrelated to effect sizes on any outcome.

In summary, we note that only diversity training versus education and publication status played a role in affective-based outcomes, the primary focus of our study. Control variables played a stronger role in cognitive-based outcomes and had little effect on skill-based outcomes.

[2]We thank an anonymous reviewer for suggesting this analysis.

DOI: 10.1002/job

Discussion

The current study contributes to our understanding of diversity training in several ways. Using prior theory and research on diversity, attitudes, and training, we examined potential differential effects of diversity training on affective-based, cognitive-based, and skill-based outcomes, and we identified features that could be incorporated into diversity training to produce stronger effects particularly on affective-based outcomes. Results from the current

Table 5. Effects of control variables.

Variable	*k*	*N*	*d*	SD_d	δ	SD_δ	95% CI	80% CV	%VAR
Affective-based outcomes									
Training or education									
Diversity training	32	3337	0.23	0.30	0.26	0.24	[0.16, 0.30]	[−0.04, 0.57]	45.02
Diversity education	12	937	0.41	0.24	0.45	0.06	[0.28, 0.54]	[0.38, 0.52]	94.85
Publication status									
Published	37	2910	0.35	0.29	0.38	0.18	[0.28, 0.42]	[0.15, 0.61]	65.70
Unpublished	7	1357	0.11	0.25	0.13	0.22	[0.00, 0.22]	[−0.16, 0.41]	34.24
Cognitive-based outcomes									
Subject matter									
Generic	7	661	0.46	0.39	0.52	0.36	[0.31, 0.61]	[0.05, 0.98]	29.57
Singular	5	239	1.17	0.39	1.28	0.24	[0.89, 1.45]	[0.97, 1.59]	67.07
Multicultural	4	256	0.46	0.62	0.54	0.63	[0.21, 0.71]	[−0.27, 1.34]	17.21
Sexual harassment	4	279	0.45	0.32	0.54	0.31	[0.21, 0.69]	[0.14, 0.93]	48.97
Combination	4	139	0.87	0.34	0.92	0.00	[0.52, 1.22]	[0.92, 0.92]	100.00
Type of diversity									
Percent women									
<40%	5	428	0.33	0.16	0.38	0.00	[0.14, 0.52]	[0.38, 0.38]	100.00
40–60%	6	272	0.68	0.41	0.85	0.36	[0.43, 0.93]	[0.39, 1.32]	54.19
>60%	9	840	0.65	0.53	0.71	0.52	[0.51, 0.79]	[0.05, 1.38]	16.75
Percent Caucasian									
<40%	2	84	1.18	0.39	1.28	0.22	[0.71, 1.65]	[1.00, 1.56]	73.29
40–60%	4	564	0.58	0.29	0.62	0.25	[0.41, 0.75]	[0.30, 0.95]	34.70
>60%	6	259	0.37	0.53	0.42	0.47	[0.12, 0.62]	[-0.19, 1.03]	34.61
Trainer race									
Caucasian	2	91	1.48	0.07	1.61	0.00	[1.01, 1.95]	[1.61, 1.61]	100.00
Non-Caucasian	2	146	0.47	0.95	0.54	1.06	[0.14, 0.80]	[−0.82, 1.91]	6.25
Self-report									
Self-report	12	770	0.77	0.48	0.85	0.43	[0.62, 0.92]	[0.29, 1.40]	29.97
Non-self-report	13	954	0.50	0.44	0.60	0.47	[0.37, 0.63]	[0.00, 1.20]	28.20
Time of evaluation									
End of training	17	967	0.71	0.24	0.82	0.56	[0.58, 0.84]	[0.10, 1.53]	24.12
Post-training ≤ 1 month	3	178	0.56	0.44	0.64	0.40	[0.26, 0.86]	[0.13, 1.15]	36.53
Post-training > 1 month	3	380	0.33	0.18	0.39	0.00	[0.13, 0.53]	[0.39, 0.39]	100.00
Study rigor									
Repeated measures	10	498	1.00	0.34	1.18	0.23	[0.81, 1.19]	[0.89, 1.47]	71.27
Training versus control (post-test only)	8	813	0.40	0.24	0.45	0.12	[0.26, 0.54]	[0.30, 0.60]	78.69
Training versus control (pre-test and post-test)	7	413	0.59	0.66	0.66	0.66	[0.39, 0.79]	[−0.18, 1.50]	16.92

Note: *k* = number of studies; *N* = total sample size across all effect sizes; *d* = the sample size weighted mean observed effect size; SD_d = the sample-size weighted standard deviation; δ = sample size weighted mean observed effect size corrected for unreliability; SD_δ = standard deviation of average corrected effect size; 95% CI = 95 percent confidence interval; 80% CV = 80 percent credibility interval; %VAR = the percent of observed variance attributable to all artifacts.

DOI: 10.1002/job

study provide researchers and practitioners with information on the relative magnitude of diversity training effects on different types of learning outcomes and under different conditions. This information can be used to guide future research and guide practitioners in designing diversity training.

In brief, our results suggested that, across all training outcomes, diversity training had a small-sized to medium-sized effect. Further, as expected, diversity training had larger effects on cognitive-based and skill-based outcomes relative to affective-based outcomes, and a number of moderators played a role in effects on affective-based outcomes. Given these results and consistent with prior reviews (e.g., Kulik & Roberson, 2008b), we concluded that diversity training can have beneficial effects and should be included in organizational training curricula. However, although we now have estimates of the size of diversity training effects in general and under different conditions, our results reveal a number of questions for future research.

Theoretical implications

Two issues are particularly important to furthering theory and research on diversity training. The issues relate to the following: (i) the multiple and potentially interacting roles of implicit and explicit processes in diversity training effects; and (ii) the nature and measurement of attitudes.

Regarding the role of implicit and explicit processes in diversity training effects, we suggest that diversity training would benefit by a more intentional consideration of implicit and explicit processes in the design and delivery of training as well as in the measurement of training outcomes. The van Knippenberg et al. (2004) model suggested that the social categorization and the information-processing/decision-making perspectives underpinning diversity theory and research can be integrated by considering their interactive effects. We extended their model to try to understand differential effects on diversity training outcomes, with a particular focus on affective-based outcomes, suggesting the following: (i) that social categorization and information-processing/decision-making routes reflect implicit and explicit processes, respectively; and (ii) that we can address processing of implicit cues by using features in diversity training that facilitate social interaction and explicit processing by using features that increase trainee motivation. A more intentional and explicit consideration of how one might employ features that enhance social interaction and trainee motivation should benefit diversity training.

Regarding the nature and measurement of attitudes, we start by noting the weaker effects observed in the current study for diversity training on affective-based outcomes, compared with effects observed on cognitive-based and skill-based outcomes. Affective-based outcomes are the most frequent outcomes evaluated in diversity training (Curtis & Dreachslin, 2008), but research results are inconsistent (Kulik & Roberson, 2008b). Certainly, changing attitudes is a primary goal of diversity training. We suggest that the weaker and mixed research results relating to affective-based outcomes say less about the effectiveness of diversity training than how researchers have examined attitudes and attitude change resulting from training. We have four comments regarding attitude change and its measurement.

Typically, researchers have examined whether post-training attitudes of trainees differ from pre-training or control group attitudes. However, a small effect of training on attitudes has multiple explanations. Training could have had no effect on attitudes. Alternatively, trainees might have entered training with favorable attitudes (explicit and implicit) toward diversity or known and reported socially desirable attitudes. Indeed, research has suggested that explicit attitudes toward diversity have become more positive since the beginning of the Civil Rights movement, but some unfavorable implicit attitudes (e.g., stereotypes) still persist (e.g., Jackson & Joshi, 2011). Unfortunately, the standard research approach is unable to disentangle these effects. That is, research would benefit by using approaches that take into account the favorableness of trainee attitudes prior to training. One would hope for greater attitude change from those trainees whose initial attitudes were less favorable. Else, one could identify a desired level of attitude favorability and determine what percentage of the trainees with less favorable attitude prior to training had a desired level of attitude favorability post-training. This is similar to Sackett and Mullen's (1993)

DOI: 10.1002/job

suggestion that training evaluation could assess the percentage of trainees who have achieved a defined level of performance as a result of training.

Second, explicit attitude measures typically are used; few studies have used implicit measures to assess attitude change resulting from diversity training (Castillo et al., 2007; Rudman et al., 2001). One solution is to use implicit measures of affective-based outcomes or, more generally, use measures that are less susceptible to socially desirable responding (e.g., implicit attitude tests and forced-choice formats). Additionally, we note that recent theorizing (e.g., Gawronski & Bodenhausen, 2006) has suggested that one can use training focused on explicit attitudes to change implicit attitudes and vice versa. Bohner and Dickel (2011) also offered a model addressing mechanisms through which explicit attitude change can influence implicit attitude change. Thus, diversity training might benefit from employing training focused on both explicit and implicit attitude change, but the measurement of attitude change resulting from training similarly needs to employ implicit as well as explicit measures.

A third reason for observed weaker effects for affective-based outcomes relates to the potential difficulty of changing attitudes versus changing knowledge or behavior. That is, attitudes may be difficult to alter to the extent that they are ingrained and inculcated over a lifetime (a debatable point given recent research showing that implicit attitudes can change rather quickly, Briñol et al., 2009). Perhaps diversity training may have stronger effects on attitudes over time as opposed to post-training. Moreover, attitude change may be preceded by and influenced by knowledge and/or behavior change. For example, an individual might learn first how to communicate with diverse others (knowledge and/or behavior change), and attitude change might follow. This is consistent with the suggestion from cognitive dissonance theory (Festinger, 1957) that enacting behavioral change might in turn lead to attitude change. Individuals can either alter their perception of the incoming information or change their attitude to fall in line with their behavior toward diverse others.

Finally, an explanation for the observed weaker effect for affective-based outcomes might be found in Golembiewski's (1986) discussion of alpha, beta, and gamma change. For instance, if on a scale of 1–7, a respondent rates her explicit attitude toward diversity as "4" in a pre-training context, then after participating in diversity training realizes that her prior attitude was closer to "2," she may rate her post-training assessment of her attitude as "4" or "5." Thus, with a re-calibration of the metric, the respondent's attitude actually changed, but the measure was not sensitive to this change. Future research should investigate the different types of change that result from diversity training.

Practical implications

The implications of our results are very appealing from a practical perspective. For instance, trainers and HR managers can use this knowledge to defend and better implement diversity training interventions within their organizations. In general, trainers and HR managers can be confident that diversity training is likely to have small-sized to medium-sized beneficial effects on affective-based, cognitive-based, and skill-based outcomes. Moreover, attention focused on the nature or type of diversity training, features of the design and/or delivery of training (including trainee characteristics), and the way one assesses the effects of training might enable practitioners to enhance the benefits of diversity training.

First, researchers (e.g., Kulik & Roberson, 2008b) have suggested that there are stronger effects for diversity training focused on dissemination (cognitive based) and skills training than on awareness (affective based). We did observe stronger effects of diversity training, in general, on cognitive-based and skill-based outcomes. However, before practitioners conclude that there are weaker effects on affective-based outcomes (e.g., attitude change), they might consider first examining training effects on implicit measures of attitude change. Also, given prior research on explicit and implicit attitude change (see Bohner & Dickel, 2011, for a review) suggesting that training focused on explicit attitude change can influence also implicit attitude change and vice versa, practitioners might consider systematically building explicit and implicit foci into training.

Second, consistent with suggestions from prior research (e.g., Kulik & Roberson, 2008a; van Knippenberg & Schippers, 2007), our results indicated significant effects for a variety of moderators relating to the design and delivery of diversity training, and in particular, features related to social interaction and trainee motivation. In

DOI: 10.1002/job

relation to social interaction, training that used interdependent tasks included active (e.g., exercises) rather than passive (e.g., lecture, video) forms of instruction and used a face to face rather than computer-based format had stronger effects on affective-based outcomes. Similarly, training of longer duration and that incorporated distributed rather than massed practice had stronger effects on affective-based outcomes. In relation to trainee motivation, we observed stronger effects when diversity training reflected features that related to trainee's perceptions that training was important and relevant, such as when managers conducted the training, when trainees were employees, and when training was conducted in field settings. Although these latter results might reflect to some extent that students might enter training with more favorable attitudes, in general, relative to employees, taken together, these results were consistent with prior research describing best practices in designing and delivering training (e.g., Goldstein & Ford, 2002; Hesketh, 1997) and in managing trainee motivation (e.g., Baldwin & Magjuka, 1997; Mathieu & Martineau, 1997).

Moreover, these results suggest that practitioners who employ better practices in designing and delivering training, including better managing pre-training characteristics (e.g., trainee motivation), will obtain greater benefits from diversity training. Although an empirical test of these notions was not possible using current research, practitioners can be guided by diversity researchers who have posited the benefits of increasing psychological safety and trust to improve learning and performance in diverse teams (e.g., Edmondson & Roloff, 2009) and reducing prejudice through increased contact under optimal conditions (e.g., performing interdependent tasks; Paluck & Green, 2009).

Third, our results suggested that control variables had a weak effect on affective-based outcomes, the primary focus of our study, although they had stronger effects on cognitive-based outcomes (e.g., diversity of the training group). This implies that issues such as the presence of a needs assessment, training choice (voluntary vs. mandatory), diversity of the group being trained, and rigor of the study design have little role in diversity training effects on affective-based outcomes. However, issues such as training choice or diversity of the trained group have either little role in training outcomes or perhaps a more complex role than can be captured from information typically reported in diversity training research to date.

Limitations

As with any study, this study had limitations. One potential limitation is that we included a heterogeneous set of studies in the meta-analysis. For instance, we treated diversity training and cross-cultural training interventions similarly. We agree that this may involve different types of samples with different motivations and reasons for participation. However, we suggest that the content and purpose of both interventions are similar enough to warrant combination. In addition, cross-cultural/intercultural training studies only accounted for a small percentage of total studies (i.e., one out of 96 effect sizes, 1 percent). Also, the exclusion of these samples did not change the overall effect size, regardless of outcome. Therefore, we felt justified in including these types of studies in the overall effect size estimates.

Some might also feel that organizational generalizability is suspect because of the inclusion of student samples and lab studies. However, we were interested in the psychological effects of diversity training on change in training outcomes as opposed to how diversity training worked within organizations per se. Also, Highhouse and Gillespie (2009) suggested that student samples do not lessen the practical importance of effect sizes. Most importantly, employee samples had larger effects on affective-based and cognitive-based outcomes than student samples, and thus, effects from student samples provided more conservative estimates of training effects.

A second potential limitation is the type of design used (i.e., experimental design vs. quasi-experimental design). There might be internal and external validity concerns, depending on the design used. With this in mind, we compared the effect sizes of experimental and quasi-experimental designs but did not detect an effect on affective-based outcomes.

A third potential limitation of this study is the sample size. Whereas some may question the size of our overall sample ($k = 96$), this study compares favorably with other recently published meta-analyses on specialized training programs (e.g., Keith & Frese, 2008; Waples, Antes, Murphy, Connelly, & Mumford, 2009). Keith and Frese's

DOI: 10.1002/job

(2008) meta-analysis on error management training had a total of 24 studies, whereas the Waples et al. (2009) meta-analysis on ethics education had a total of 25 studies. Nevertheless, because of limited data (e.g., two or three studies) for some moderators, some of our moderator analyses may suffer from second-order sampling error. This occurs when the meta-analysis does not average out the primary sampling error. Hunter and Schmidt (2004) observed that $\text{Avg}(d) = \text{Avg}(\delta) + \text{Avg}(e)$. In other words, the observed meta-analytic effect size is determined by the population effect size plus the average sampling error. If the sample is large enough, $\text{Avg}(e) = 0$. $\text{Avg}(e)$ becomes non-zero when the number of studies is small and the observed $\text{Avg}(d)$ deviates from the population effect size. However, most of our moderation analyses use a greater number of studies than the example given.

Future research

As discussed in theoretical implications, future research integrating attitude and training theory and models into diversity training research is needed. For example, we need to know more about how to design and deliver training that aids trainees in having more adaptive reactions and responses to diversity attributes through both implicit and explicit processes. Extending the van Knippenberg et al. (2004) model, we suggest that we need to know more about how features relating to social interaction might be used to help trainees process implicit cues, that is, to process social categorizations in a beneficial way. We also need to know more about how features relating to trainee motivation might be used to help trainees get the most out of explicit processing of diversity attributes. For example, we were able to code only indicators of trainee motivation (e.g., trainer source), and more direct measures of motivation or perceptions of training importance and value are needed.

Also, as noted in the introduction, diversity training is a component of a larger set of diversity initiatives within the organization. Researchers have observed beneficial effects of diversity training on proximal outcomes (e.g., Kulik & Roberson, 2008b). However, although researchers have begun to look at issues of transfer (Holladay et al., 2008; Roberson et al., 2001), additional research examining aspects of the work environment that foster greater transfer gains from diversity training is needed. Also, research linking proximal outcomes with distal outcomes is needed (e.g., work performance in diverse teams, Edmondson & Roloff, 2009). Finally, research examining the role of training in relation to organizational diversity policies, initiatives, and distal goals is needed (e.g., minority hiring).

Further, future research examining diversity training effects on affective-based outcomes that takes into account trainees' attitudes prior to training and perhaps using other approaches to assessing training outcomes is needed (e.g., percent of trainees with a desired level of favorable attitudes at the end of training). For example, one would expect little attitude change from trainees with initially favorable attitudes. We need to know more about who those trainees are, how they react to training, and how they might influence attitude change in other trainees.

Conclusion

Organizations have been citing the "business case" for diversity since the 1980s (e.g., Bendick et al., 2001), but it has been difficult to obtain the benefits of diversity (Edmondson & Roloff, 2009). Diversity training is a key component of diversity initiatives, demonstrating stronger effects on cognitive-based and skill-based outcomes than on affective-based outcomes. However, there has been little empirical evidence to suggest when and under what conditions diversity training is most beneficial. Using theory and research on diversity, attitudes, and training, we examined potential differential effects on affective-based, cognitive-based, and skill-based outcomes and examined potential moderators of those effects with a focus on affective-based outcomes. Using meta-analytic techniques, we provided evidence of the beneficial effects of diversity training on multiple outcomes and how moderators influenced affective-based outcomes. Our results provide a foundation that can help researchers and practitioners build a case for diversity management in organizations, frame future research to enrich our understanding of diversity training effects, and aid practitioners in designing more effective diversity training.

DOI: 10.1002/job

Author biographies

Zachary T. Kalinoski is an associate scientist at Aptima, Inc. He received his PhD in Industrial and Organizational Psychology from Wright State University. His research interests include training and development, learning from errors, business ethics, work motivation, and diversity/cross-cultural issues. His applied work has focused on examining skills, competencies, behaviors, contextual attributes, and team states in cross-cultural settings, evaluating applied data with qualitative and quantitative statistical methods, designing performance assessment metrics, using job analytic data to facilitate training gap analysis, and designing training and validating team-level constructs in virtual worlds.

Debra Steele-Johnson is an associate professor of Psychology in the Industrial and Organizational Psychology PhD concentration at Wright State University. She received her PhD in Industrial and Organizational Psychology from the University of Minnesota. Her research interests include training and motivation for individuals and teams. She has published over 25 peer-reviewed journal articles and book chapters. Her work has appeared in numerous scholarly journals, including *Journal of Applied Psychology, Human Performance, Journal of Applied and Social Psychology, Journal of Applied Behavioral Science,* and *International Journal of Selection and Assessment*.

Elizabeth J. Peyton completed her MS and is currently completing her PhD in Industrial and Organizational Psychology at Wright State University. Her research focuses on how attitudes influence leadership and performance in organizations. She has applied experience also in marketing research, working for Interbrand and Ipsos.

Keith A. Leas completed his MS in Industrial and Organizational Psychology at Wright State University. His research interests include training, organizational development, and human-computer interaction. He has worked with several companies in designing and implementing training programs and organizational change projects. He currently works for Caterpillar, Inc.

Julie Steinke completed her MS and is currently completing her PhD in Industrial and Organizational Psychology at Wright State University. She received an MS in Applied Sport Science/Sport Psychology from Indiana University. She is a lecturer in the Department of Psychological Sciences at Northern Kentucky University. Her research interests include motivation, leadership, and teams, focusing on cooperation, competition, and conflict among team members. Her applied work has focused on astronaut selection (working at the National Aeronautics and Space Administration), organizational assessments and reorganization projects for non-profit agencies, and training and development within college athletics.

Nathan A. Bowling is a professor of Psychology in the Industrial and Organizational Psychology PhD concentration at Wright State University. He received his PhD in Industrial and Organizational Psychology from Central Michigan University. His research interests include job satisfaction, counterproductive work behavior, and occupational stress. He has published over 40 peer-reviewed journal articles and book chapters. His work has appeared in several top scholarly journals, including the *Journal of Applied Psychology, Journal of Occupational Health Psychology,* and *Journal of Occupational and Organizational Psychology*.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, *88*, 234–245.

*Badhesha, R. S., Schmidtke, J. M., Cummings, A., & Moore, S. D. (2008). The effects of diversity training on specific and general attitudes toward diversity. *Multicultural Education & Technology Journal*, *2*, 87–106.

*Bailey, A., Barr, O., & Bunting, B. (2001). Police attitudes toward people with intellectual disability: An evaluation of awareness training. *Journal of Intellectual Disability Research*, *45*, 344–350.

Baldwin, T. T., & Magjuka, R. J. (1997). Training as an organizational episode: Pretraining influences on trainee motivation. In J. K. Ford, S. J. Kozlowski, K. Kraiger, & M. S. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 99–128). Mahwah, NJ: Erlbaum.

*Barker, S. A. (2004). Improving interpersonal job skills by applying cross-cultural training. *Journal of European Industrial Training*, *28*, 215–233.

DOI: 10.1002/job

Bendick, M., Egan, M. L., & Lofhjelm, S. M. (2001). Workforce diversity training: From anti-discrimination compliance to organizational development. *Human Resource Planning*, *24*, 10–25.

Bennett, J. (2006). Achieving race equality through training: A review of approaches in the UK. *The Journal of Mental Health Workforce Development*, *1*, 5–11.

*Blakely, G. L., Blakely, E. H., & Moorman, R. H. (1998). The effects of training on perceptions of sexual harassment allegations. *Journal of Applied Social Psychology*, *28*, 71–83.

Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology*, *62*, 391–417.

*Brathwaite, A. E. C. (2005). Evaluation of a cultural competence course. *Journal of Transcultural Nursing 16*(4), 361–369.

Briñol, P., Petty, R. E., & McCaslin, M. J. (2009). Changing attitudes on implicit versus explicit measures. In R. E. Petty, R. H. Fazio, P. Briñol (Eds.), *Attitudes. Insights from the new implicit measures* (pp. 285–326). New York: Psychology Press.

*Brown, E. L. (2004). What precipitates change in cultural diversity awareness during a multicultural course: The message or the method? *Journal of Teacher Education*, *55*, 325–340.

*Bush, V. D., & Ingram, T. N. (2001). Building and assessing cultural diversity skills: implications for sales training. *Industrial Marketing Management*, *30*, 65–76.

*Byrnes, D. A., & Kiger, G. (1990). The effect of a prejudice-reduction simulation on attitude change. *Journal of Applied Social Psychology*, *20*, 341–356.

*Castillo, L. G., Brossart, D. F., Reyes, C. J., Conoley, C. W., & Phoummarath, M. J. (2007). The influence of multicultural training on perceived multicultural counseling competencies and implicit racial prejudice. *Journal of Multicultural Counseling and Development*, *35*, 243–254.

*Celik, H., Abma, T. A., Klinge, I., & Widdershoven, G. A. M. (2011). Process evaluation of a diversity training programme: The value of a mixed method strategy. *Evaluation and Program Planning*, *35*, 54–65.

Chavez, C. I., & Weisinger, J. Y. (2008). Beyond diversity training: A social infusion of cultural inclusion. *Human Resource Management*, *47*, 331–350.

*Chiasson, J. (2006). Lifting the veil of heterosexism: Effecting attitudinal change towards sexual minorities (Unpublished doctoral dissertation). Claremont Graduate University, Claremont, CA.

*Combs, G. M., & Luthans, F. (2007). Diversity training: Analysis of the impact of self-efficacy. *Human Resource Development Quarterly*, *18*, 91–120.

*Cornett-DeVito, M. M., & McGlone, E. L. (2000). Multicultural communication training for law enforcement officers: A case study. *Criminal Justice Policy Review*, *11*, 234–253.

Curtis, E. F., & Dreachslin, J. L. (2008). Diversity management interventions and organizational performance. A synthesis of current literature. *Human Resource Development Review*, *7*(1), 107–134.

*Davis, K. J. (2008). Workforce diversity training programs: Non-convergence in cognitive structures. *International Journal of Business Research*, *8*, 38–51.

*De Meuse, K. P., Hostager, T. J., & O'Neill, K. S. (2007). A longitudinal evaluation of senior managers' perceptions and attitudes of a workplace diversity training program. *Human Resource Planning*, *30*, 38–46.

*Dean, S. L. (2008). How openness to experience and prejudicial attitudes shape diversity training outcomes (Unpublished master's thesis). Wright State University, Dayton, OH.

Deutsch, M. (1949). A theory of co-operation and competition. *Human Relations*, *2*, 129–152.

*Dickson, G. L., Argus-Calvo, B., & Tafoya, N. G. (2010). Multicultural counselor training experiences: Training effects and perceptions of training among a sample of predominantly Hispanic students. *Counselor Education and Supervision*, *49*, 247–265.

*Doorenbos, A. Z., & Schim, S. M. (2004). Cultural competence in hospice. *The American Journal of Hospice & Palliative Care*, *21*, 28–32.

*Edelstein, S. (2007). Measurement of diversity tolerance of nutrition professionals using the Tolerance of Ambiguity Scale. *Topics in Clinical Nutrition*, *22*, 130–136.

Edmondson, A. C., & Roloff, K. S. (2009). Overcoming barriers to collaboration: Psychological safety and learning in diverse teams. In E. Salas, G. F. Goodwin, & C. S. Burke (Eds.), *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches* (pp. 183–208). New York, NY: Routledge.

Engberg, M. E. (2004). Improving intergroup relations in higher education: A critical examination of the influence of educational interventions on racial bias. *Review of Educational Research*, *74*(4), 473–524.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

*Flavin, C. (1997). Cross-cultural training for nurses: A research-based education project. *The American Journal of Hospice & Palliative Care*, *14*, 121–126.

*Gany, F., & de Bocanegra, H. T. (1996). Maternal–child immigrant heath training: Changing knowledge and attitudes to improve health care delivery. *Patient Education and Counseling*, *27*, 23–31.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Goldstein, I. L., & Ford J. K. (2002). *Training in organizations: Needs assessment, development, and evaluation*. Belmont, CA: Wadsworth.

DOI: 10.1002/job

Golembiewski, R. T. (1986). Contours in social change: Elemental graphics and a surrogate variable for gamma change. *Academy of Management Review*, *11*, 550–566.

*Gulick, L., Jose, I., Peddie, C., King, E., Kravitz, D., & Ferro, G. (2009). Enhancing diversity training outcomes: How recognizing bias about bias will impact diversity-supportive behaviors and hiring decisions. Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, New Orleans, LA.

Gutiérrez, L., Kruzich, J., Jones, T., & Coronado, N. (2000). Identifying goals and outcome measures for diversity training: A multi-dimensional framework for decision-makers. *Administration in Social Work*, *24*, 53–70.

*Guy-Walls, P. (2007). Exploring cultural competence practice in undergraduate social work education. *Education*, *127*, 569–580.

*Hanover, J. M. B., & Cellar, D. F. (1998). Environmental factors and the effectiveness of workforce diversity training. *Human Resource Development Quarterly*, *9*, 105–124.

*Hauenstein, N. M. A., Findlay, R. A., & McDonald, D. P. (2010). Using situational judgment tests to assess training effectiveness: Lessons learned evaluating military equal opportunity advisor trainees. *Military Psychology*, *22*, 262–281.

*Hayes, S. C., Bissett, R., Roget, N., Padilla, M., Kohlenberg, B. S., Fisher, G., . . . Niccolls, R. (2004). The impact of acceptance and commitment training and multicultural training on the stigmatizing attitudes and professional burnout of substance abuse counselors. *Behavior Therapy*, *35*, 821–835.

Hesketh, B. (1997). Dilemmas in training for transfer and retention. *Applied Psychology: An International Review*, *46*, 317–339.

Highhouse, S., & Gillespie, J. Z. (2009). Do samples really matter that much? In C. E. Lance, & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences* (pp. 247–266). New York: Routledge.

*Hill, M. E., & Augoustinos, M. (2001). Stereotype change and prejudice reduction: Short- and long-term evaluation of a cross-cultural awareness programme. *Journal of Community & Applied Social Psychology*, *11*, 243–262.

Hite, L. M., & McDonald, K. S. (2006). Diversity training pitfalls and possibilities: An exploration of small and mid-size U.S. organizations. *Human Resource Development International*, *9*, 365–377.

*Holladay, C. L., Anderson, D., Gilbert, S., & Turner, S. (2008). Evaluating diversity training effectiveness: Self-efficacy as an enabler of transfer. Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, San Francisco, CA.

Holladay, C. L., Knight, J. L., Paige, D. L., & Quiñones, M. A. (2003). The influence of framing on attitudes toward diversity training. *Human Resource Development Quarterly*, *14*, 245–263.

Holladay, C. L., & Quiñones, M. A. (2005). Reactions to diversity training: An international comparison. *Human Resource Development Quarterly*, *16*, 529–544.

Holladay, C. L., & Quiñones, M. A. (2008). The influence of training focus and trainer characteristics on diversity training effectiveness. *The Academy of Management Learning and Education*, *7*, 343–354.

Hollister, L. E., Day, N. E., & Jesaitis, P. T. (1993). Diversity programs: Key to competitiveness or just another fad? *Organizational Development Journal*, *11*, 49–59.

*Hostager, T. J., & de Meuse, K. P. (2008). The effects of a diversity learning experience on positive and negative diversity perceptions. *Journal of Business and Psychology*, *23*, 127–139.

Huffcutt, A. I., & Arthur, W., Jr. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, *80*, 327–334.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting for bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage.

Jackson, S. E., & Joshi, A. (2011). Work team diversity. In S. Zedeck (Ed.), *APA handbook of industrial and organizational psychology: Building and developing the organization* (Vol. 1, pp. 651–686). Washington, DC: American Psychological Association.

*Jefferson, M. R. (2001). The influence of diversity training on the attitudes and perceptions of freshmen students (Unpublished doctoral dissertation). Seton Hall University, South Orange, NJ.

Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, *71*(4), 589–617.

Kath, L. M. (2004). How cynicism about organizational change can influence the effectiveness of socialization training: Sexual harassment training awareness as exemplar (Unpublished doctoral dissertation). University of Connecticut, Storrs, CT.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology*, *93*(1), 59–69.

King, E. B., Gulick, L. M. V., & Avery, D. R. (2010). The divide between diversity training and diversity education: Integrating best practices. *Journal of Management Education*, *34*(6), 891–906.

van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology*, *89*, 1008–1022.

van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology*, *58*, 515–541.

DOI: 10.1002/job

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*, 311–328.

*Krings, F., Bollmann, G., & Palazzo, B. (2009). Diversity "spielerisch" trainieren: Auswirkungen auf Einstellungen und Sensibilität gegenüber Diversity bei Führungskräften. *Zeitschrift für Arbeits- und Organisationspsychologie*, *53*, 33–38.

*Kulik, C. T., Pepper, M. B., Roberson, L., & Parker, S. K. (2007). The rich get richer: Predicting participation in voluntary diversity training. *Journal of Organizational Behavior*, *28*, 753–769.

Kulik, C. T., & Roberson, L. (2008a). Common goals and golden opportunities: Evaluations of diversity education in academic and organizational settings. *The Academy of Management Learning and Education*, *7*(3), 309–331.

Kulik, C. T., & Roberson, L. (2008b). Diversity initiative effectiveness: What organizations can (and cannot) expect from diversity recruitment, diversity training, and formal mentoring programs. In A. P. Brief (Ed.), *Diversity at work* (pp. 265–317). Cambridge: Cambridge University Press.

*Landis, D., Brislin, R. W., & Hulgus, J. F. (1985). Attributional training versus contact in acculturative learning: A laboratory study. *Journal of Applied Social Psychology*, *15*, 466–482.

Liberman, B. E., Block, C. J., & Uyebuko S. M. (2010). The role of individuating information on perceived diversity trainer effectiveness. Paper presented at the annual meeting of the Society of Industrial and Organizational Psychology, Atlanta, GA.

*Luger, L. (2011). Enhancing cultural competence in staff working with people with drug and alcohol problems: A multidimensional approach to evaluating the impact of education. *Social Work Education*, *30*, 223–235.

Lynch, F. R. (1997). *The diversity machine: The drive to change the "white male workplace*." New York: Free Press.

*Majumdar, B., Keystone, J. S., & Cuttress, L. A. (1999). Cultural sensitivity training among foreign medical graduates. *Medical Education*, *33*, 177–184.

*Mak, A. S., & Buckingham, K. (2007). Beyond communication courses: Are there benefits in adding skills-based Excell™ sociocultural training? *International Journal of Intercultural Relations*, *31*, 277–291.

Martins, L. L., Milliken, F. J., Wiesenfeld, B. M., & Salgado, S. R. (2003). Racioethnic diversity and group members' experiences. *Group and Organization Management*, *28*(1), 75–106.

Mathieu, J. E., & Martineau, J. W. (1997). Individual and situational influences in training motivation. In J. K. Ford, S. J. Kozlowski, K. Kraiger, & M. S. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 193–222). Mahwah, NJ: Erlbaum.

*Mausehund, J. A., Timm, S. A., & King, A. S. (1995). Diversity training: Effects of an intervention treatment on nonverbal awareness. *Business Communication Quarterly*, *58*, 27–30.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *Academy of Management Review*, *21*(2), 402–433.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105–125.

*Moyer, R. S., & Nath, A. (1998). Some effects of brief training interventions on perceptions of sexual harassment. *Journal of Applied Social Psychology*, *28*, 333–356.

Nemetz, P. L., & Christiansen, S. L. (1996). The challenge of cultural diversity: Harnessing a diversity of views to understand multiculturalism. *Academy of Management Review*, *21*, 434–462.

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

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Paluck, E. L. (2006). Diversity training and intergroup contact: A call to action research. *Journal of Social Issues*, *62*, 577–595.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A review and assessment of research and practice. *Annual Review of Psychology*, *60*, 339–367.

*Parker, W. M., Moore, M. A., & Neimeyer, G. J. (1998). Altering white racial identity and interracial comfort through multicultural training. *Journal of Counseling and Development*, *76*, 302–310.

*Pedersen, P. J. (2010). Assessing intercultural effectiveness outcomes in a year-long study abroad program. *International Journal of Intercultural Relations*, *34*, 70–80.

Pendry, L. F., Driscoll, D. M., & Field, S. C. T. (2007). Diversity training: Putting theory into practice. *Journal of Occupational and Organizational Psychology*, *80*, 27–50.

Perry, E. L., Kulik, C. T., & Field, M. P. (2009). Sexual harassment training: Recommendations to address gaps between the practitioner and research literatures. *Human Resource Management*, *48*, 817–837.

*Perry, E. L., Kulik, C. T., & Schmidtke, J. (1998). Individual differences in the effectiveness of sexual harassment awareness training. *Journal of Applied Social Psychology*, *28*, 698–723.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*, 751–783.

DOI: 10.1002/job

*Preusser, M. K., Bartels, L. K., & Nordstrom, C. R. (2011). Sexual harassment training: Person versus machine. *Public Personnel Management*, *40*, 47–61.
*Probst, T. M. (2003). Changing attitudes over time: Assessing the effectiveness of a workplace diversity course. *Teaching of Psychology*, *30*, 236–239.
*Pruegger, V. J., & Rogers, T. B. (1994). Cross-cultural sensitivity training: Methods and assessment. *International Journal of Intercultural Relations*, *18*, 369–387.
*Randolph, G., Landis, D., & Tzeng, O. C. S. (1977). The effects of time and practice upon culture assimilator training. *International Journal of Intercultural Relations*, *1*, 105–119.
*Reynolds, L. (2010). Aging and disability awareness training for drivers of a metropolitan taxi company. *Activities, Adaptation, & Aging*, *34*, 17–29.
*Robb, L. A., & Doverspike, D. (2001). Self-reported proclivity to harass as a moderator of the effectiveness of sexual harassment-prevention training. *Psychological Reports*, *88*, 85–88.
Roberson, L., Kulik, C. T., & Pepper, M. B. (2001). Designing effective diversity training: Influence of group composition and trainee experience. *Journal of Organizational Behavior*, *22*, 871–885.
*Robinson, B., & Bradley, L. J. (1997). Multicultural training for undergraduates: Developing knowledge and awareness. *Journal of Multicultural Counseling and Development*, *25*, 281–289.
*Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, *81*, 856–868.
Rynes, S., & Rosen, B. (1995). A field survey of factors affecting the adoption and perceived success of diversity training. *Personnel Psychology*, *48*, 247–270.
Sackett, P. R., & Mullen, E. J. (1993). Beyond formal experimental design: Towards an expanded view of the training evaluation process. *Personnel Psychology*, *46*, 613–627.
*Sanchez, J. L., & Medkik, N. (2004). The effects of diversity awareness training on differential treatment. *Group and Organization Management*, *29*, 517–536.
*Scher, L. J. (2008). Beyond acceptance: An evaluation of the Safe Zone Project in a clinical psychology doctoral program (Unpublished doctoral dissertation). Long Island University, C. W. Post Campus, Roslyn Heights, NY.
*Soble, J. R., Spanierman, L. B., & Liao, H.-Y. (2011). Effects of a brief video intervention on White university students' racial attitudes. *Journal of Counseling Psychology*, *58*, 151–157.
Steel, P., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology*, *87*(1), 96–111.
*Steinfeldt, J. A., & Wong, Y. J. (2010). Multicultural training on American Indian issues: Testing the effectiveness of an intervention to change attitudes toward native-themed mascots. *Cultural Diversity and Ethnic Minority Psychology*, *16*, 110–115.
*Stewart, T. L., LaDuke, J. R., Bracht, C., Sweet, B. A. M., & Gamarel, K. E. (2003). Do the eyes have it? A program evaluation of Jane Elliott's "Blue Eyes/Brown Eyes" diversity training exercise. *Journal of Applied Social Psychology*, *33*, 1898–1921.
Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.
*Tang, T. S., Hernandez, E. J., & Adams, B. S. (2004). "Learning by teaching": A peer-teaching model for diversity training in medical school. *Teaching and Learning in Medicine*, *16*, 60–63.
*Thomas, C. H. (2008). An analysis of the effects of diversity training on employment practices (Unpublished doctoral dissertation). University of Phoenix, Phoenix, AZ.
*Thomas, V. J., & Cohn, T. (2006). Communication skills and cultural awareness courses for healthcare professionals who care for patients with sickle cell disease. *Issues and Innovations in Nursing Education*, *47*, 480–488.
Waples, E. P., Antes, A. L., Murphy, S. T., Connelly, S., & Mumford, M. D. (2009). A meta-analytic investigation of business ethics instruction. *Journal of Business Ethics*, *87*, 133–151.
Wiethoff, C. (2004). Motivation to learn and diversity training: Application of the theory of planned behavior. *Human Resource Development Quarterly*, *11*, 35–60.
*Williams, C. C. (2005). Training for cultural competence: Individual and group processes. *Journal of Ethnic and Cultural Diversity in Social Work*, *14*, 111–143.
Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior*, *20*, 77–140.
*Zarubin, A. (2008). Effect of attitudes on results of diversity training (Unpublished master's thesis). Bowling Green State University, Bowling Green, OH.

DOI: 10.1002/job