Meta-Analytic Evaluations of Interventions to Improve Ethnic Attitudes

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich Psychologie

der Philipps-Universität Marburg

vorgelegt

von

Dipl.-Psych. Gunnar Lemmer

aus Gießen

Marburg/Lahn 2011

Vom Fachbereich Psychologie der Philipps-Universität Marburg als Dissertation

am 07.10.2011 angenommen.

Erstgutachter: Prof. Dr. Ulrich Wagner (Philipps-Universität Marburg)

Zweitgutachter: Prof. Dr. Mario Gollwitzer (Philipps-Universität Marburg)

Tag der mündlichen Prüfung am 15.10.2011

In Liebe und Dankbarkeit für Erich (gest. 15.4.2010) und Annelen Lemmer

Table of Contents

1. Introduction

1. The Phenomenon: Ethnic Prejudice.

Interethnic tensions are (still) a fact of social live. Many regions (e.g., the Middle East) are marked by enduring—termed as intractable—conflicts between ethnic groups. Even in areas where such conflicts do not or no longer exist (e.g., USA, Western Europe), ethnic discrimination (i.e., biased behavior that is solely based on ethnic group memberships) is an undesired reality. Studies demonstrate that people with a different ethnic background (e.g., persons with Turkish roots [in Germany]; African Americans [in the USA]) are disadvantaged in a huge variety of domains such as daily life activities (e.g., Klink & Wagner, 1999), labor markets (e.g., Carlsson & Rooth, 2007; Kaas & Manger, in press), and housing markets (e.g., Ahmed & Hammersted, 2008; Riach & Rich 2002; Massey & Lundy, 2001). In addition, severe acts of group-motivated violence are documented (e.g., European Union Agency for Fundamental Rights, 2010; U.S. Department of Justice, 2009; U.S. Department of State, 2009).

A substantial body of research (e.g., Bushman & Bonacci, 2004; Dovidio, Gaertner, Nier, Kawakami, & Hodson, 2004; Dovidio, Kawakami, & Gaertner, 2002; Schütz & Six, 1996) supports the nearby assumption that *ethnic prejudice* is a central fundament of these types of negative intergroup behavior. Gordan Allport (1954) defines prejudice in his influential publication *The Nature of Prejudice* as an "antipathy based on faulty and inflexible generalization" (p. 9). He adds that this antipathy "may be felt or expressed . . . may be directed toward a group as a whole, or toward an individual because he is a member of that group" (p. 9). Nowadays, the phenomenon of prejudice is typically considered as a negative group-based attitude toward an individual member of an outgroup or toward an outgroup as a whole. As other attitudes (see Eagly & Chaiken, 1993), prejudice encompasses three components: cognitive, affective, and conative. The cognitive facet of prejudice focuses on specific thoughts or beliefs about an outgroup or its members (e.g., "Blacks are getting too

demanding in their push for equal rights"). In contrast, the affective component refers to emotions toward outgroups or members of them (e.g., dislike). Finally, the conative facet contains associations with past or intended behavior toward an outgroup or its members. In addition to this explicit and content-based conceptualization of (intergroup) attitudes, further approaches, for instance, also consider attitude strength (e.g., Krosnick & Petty, 1995) as well as implicit attitudes (e.g., Fazio & Olson, 2003), that is, attitudes that are generally unacknowledged or out of awareness.

Research shows that ethnic prejudice not only exists in regions with intractable conflicts but is also (still) a widespread phenomenon in other areas (e.g., Abrams & Houston, 2006; Holmes, Murachver, & Bayard, 2001; Schweitzer, Perkulidis, Krome, Ludlow, & Ryan, 2005; Zick, Pettigrew & Wagner, 2008). Therefore, the issue arises: What are the reasons for prejudice and what can be done to diminish it?

Since the middle of the 20th century, social psychology and associated disciplines provide a growing body of theoretical approaches that are able to explain the existence of negative intergroup attitudes. These approaches are supported by empirical evidence and have identified prejudice-promoting aspects such as authoritarianism (Altemeyer, 1996), social dominance orientation (Pratto, Sidanius Stallworth, & Malle, 1994; Sidanius & Pretto, 1999), social categorizations and identity processes (Tajfel & Turner, 1979, 1986), perceived realistic group conflicts (see Sherif, 1966; Sherif & Sherif, 1969), relative deprivation (Gurr, 1969), as well as intergroup threats (Stephan & Stephan, 2000). Interestingly, among the approaches with direct relevance to intergroup attitudes, the not yet mentioned intergroup contact theory (Allport, 1954; Pettigrew & Tropp, 2011) has conceptually—together with several categorization models that have been developed in the context of intergroup contact (see e.g., Brewer & Miller, 1984; Gaertner & Dovidio, 2000; Pettigrew, 1998)—an

exceptional position: It deals with the social problem of prejudice from the other, positive side. According to the contact approach, prejudice can be reduced by intergroup interactions.

In addition to the existence of theories that help to understand the phenomenon of ethnic prejudice, concrete knowledge about interventions that are able to improve ethnic attitudes and to diminish ethnic tensions is of major importance.

1.2. Interventions to Improve Ethnic Attitudes

As documented by reviews (e.g., Denson, 2009; Dessel & Rogge, 2008; Engberg, 2004; McGregor, 1989, 1993; Paluck & Green, 2009), a multiplicity of interventions have been implemented in order to improve ethnic attitudes. A premise for a fruitful test of the effectiveness of these programs is a taxonomy that is able to adequately classify them. With such a typology, a comparison of their outcomes and an identification of the most beneficial intervention types would be possible. Unfortunately, as described in detail in Manuscript #2, previous suggestions for the classification of prejudice reduction strategies are marked by essential shortcomings. We have therefore conceptualized an alternative way to classify initiatives to improve ethnic attitudes.

At the most general level, contact programs are differentiated from information interventions. Contact initiatives are interventions that are explicitly based on the intergroup contact theory and bring people with a different ethnic background in direct (i.e., face-to-face) contact or let the participants experience indirect contact (i.e., variants of contact that are not based on face-to-face interactions). In contrast, the strategies assigned to the broad category of information programs give some kind of information-based but not contact-associated input in order to improve ethnic attitudes. Although many of these programs are not explicitly based on one of the above mentioned (or other) theories, they are typically associated with the mechanisms suggested by them (see Manuscript #2).

At a more fine-grained level, we suggest to differ between two types of direct contact (i.e., contact meeting and cooperative learning) and three types of indirect contact (i.e., extended, virtual, and imagined) interventions. In reference to information programs, we introduce a multi-axial taxonomy that encompasses the content (axis 1: knowledge, empathy, social-cognitive skills, and combinations thereof), method (axis 2: passive, active, as well as passive and active), and duration (axis 3: one-day vs. multi-day; number of net intervention hours) of information strategies. Details regarding the respective interventions and program characteristics are provided in Manuscript #1 (contact interventions) and Manuscript #2 (information interventions).

The essential questions that are addressed in this dissertation are: Can ethnic prejudice typically be reduced by the implementation of these programs and, in this context, which types of interventions are the most effective and can be recommended for future realizations—at least from the perspective of their effectiveness. Although the types of programs just mentioned are explicitly based on a well supported theory (in the case of contact interventions) or associated with aspects that have been empirically demonstrated to be central for intergroup prejudice (in the case of information interventions), their actual effectiveness has to be scrutinized by means of outcome evaluations.

1.3. Outcome evaluation – An Instrument to Assess the Effect of an Intervention

To evaluate an entity generally means to assess it in regard to one or more criteria. In this context, Suchman (1967) coined the term evaluation research. This label is confusing, it does not refer to research about evaluation (methods), but to evaluations that are conducted in consideration of scientific criteria and with the utilization of scientific research methods. Such evaluations can be classified on several axes of differentiation, for instance, regarding the type of entity that is evaluated (e.g., persons, laws, social programs), the evaluation

criterion (e.g., implementation quality, outcome, i.e., the effectiveness), and the applied research method (e.g., naturalistic observation, experimental or quasi-experimental design).

The work presented in this dissertation considers evaluations of the outcome of interventions to improve ethnic attitudes with experimental or quasi-experimental designs. I use the term *outcome evaluation* to characterize this type of research. An outcome evaluation refers to "the assessment of the effects of interventions upon the populations they are intended to benefit" (Lipsey & Cordray, 2000, p. 2).

Outcome evaluations should ensure that the outcome is really the consequence of the implemented program. Therefore, they have to be designed in a way that eliminates alternative explanations for the (possible) observed effects. Regarding this aspect, Donald Campbell and Julian Stanley (1966; see also Campbell, 1957, Campbell & Stanley, 1961) published the influential volume *Experimental and Quasi-experimental Designs for Research*. The authors introduced an explicit distinction between the internal and external validity of causal statements (two further concepts were added later on; see Cook & Campbell, 1976; Shadish, Cook, & Campbell, 2002). While external validity refers to the amount the findings are generalizable to other persons, settings, and time periods, internal validity deals with the degree the assumption of a cause (here: intervention) effect (here: improvement of ethnic attitudes) relation is justifiable. The authors also provide a systematization of potential threats to internal validity (e.g., maturation, history, i.e., the occurrence of specific events between the pre- and the posttest) as well as of several experimental and quasi-experimental designs that are more or less able to control for threats to internal validity.

Among the presented designs, some are typically used in outcome evaluations of social programs and are of relevance for the work conducted in the context of this dissertation. These designs are configurations of the following components: pretest, control group, random assignment of participants to the experimental or the control group, and posttest.

The first design (labeled as *posttest only with control* in Manuscript #1 and Manuscript # 2) contains a posttest, a control group, and randomization of participants to groups. In contrast, a pretest is not included.

The second design is similar to the first one but lacks a random assignment of individuals.

The third one (termed as *pretest-posttest with control* in the two manuscripts) includes a pretest, a control group, random assignment of participants, and a posttest.

The fourth design resembles the third one but individuals are not randomly allocated to conditions.

Finally, the fifth design (labeled as *pretest-posttest single group* in the following manuscripts) does not contain a control group but includes a pretest as well as a posttest.

Some aspects are important to note. First, We decided to not consider evaluations that utilized the second design in the meta-analyses that are presented in Manuscript #1 and Manuscript #2. This design is—due to a lack of randomization and the absence of a pretest—subject to essential threats to internal validity. In addition, only a small number of studies that satisfied our other inclusion criteria regarding type of intervention, dependent variable(s) and formal characteristics (see below for more information) used this design. Second, we term both the third and the fourth design as pretest-posttest with control and included both of them in the two meta-analyses. We did so because random assignment of participants to conditions is often not possible in the field of outcome evaluations of social interventions and because the usage of a pretest offers the possibility to control for pre-interventional differences between the intervention and the control group in the dependent variable(s)—usually the most serious confounder. Third, we also considered studies in the presented meta-analyses that used the fifth design but analyzed them separately from research that realized a more elaborated design. Although the pretest-posttest single group design is marked by essential

threats to internal validity (e.g., history and maturation), we decided to not ignore the corresponding work. This decision is based on the fact that a multiplicity of the studies that were found by us with initial searches evaluated often carefully designed and implemented interventions but had to use this restricted design because an adequate control group was not available.

In summary, a single outcome evaluation that is considered in my work can test the effectiveness of a prejudice reduction program with the use of a posttest only with control design, a pretest-posttest with control design, or a pretest-posttest single group design. Additionally, it can—albeit in a more or less compelling way—either provide evidence for the positive impact of an intervention (i.e., when the treatment condition outperforms the control group or the pre-test control condition) or indicate the absence of a positive effect.

However, if one is interested in answering the question whether interventions to reduce ethnic prejudice are generally effective and in identifying strategies that are the most promising, an individual outcome evaluation is of very limited use. An obvious possibility that seems to enable more general conclusions is to take a quick glance at the results of several well-known, often cited studies and/or at work that is immediately found with internet searches. This technique is highly problematic since the conclusion is based on an unrepresentative and possibly biased sample (see Rothstein, Sutton, & Borenstein, 2005). The sample could, for instance, only include studies that are well-known because they have especially advantageous and/or surprising results. In addition, the detection of sub-types of interventions that are particularly beneficial (i.e., the identification of one or multiple moderator variables) is hardly possible with this cursory procedure. Therefore, the use of specialized methods that allow to summarize previous literature in an exhaustive, structured, and unbiased manner is indicated.

1.4. Meta-Analysis – An Instrument to Summarize Outcome Evaluations

In reference to Gene Glass (1976), syntheses that integrate the results of multiple studies can be methodologically classified as a third type of scientific data acquisition. He differentiates it from primary research (i.e., collecting new data to analyze a research question) and secondary research (i.e., re-analyzing the data of a given study with new methods and/or further research questions).

Considering this third type of research, three approaches can be differentiated: narrative review, quantitative integration based on significance levels, and quantitative integration based on effect sizes. In the context of narrative (or qualitative) reviews, literature that deals with the respective research question is searched, reviewed, and verbally summarized. Although this instrument is often used, it is marked by several shortcomings (see Rosenthal & DiMatteo, 2001) such as deficits concerning structure and transparency, subjective biases regarding the selection and weighting of studies, and problems with respect to the identification of moderators.

A second set of research syntheses refers to the quantitative integration of significance levels via the use of the vote counting technique (Light & Smith, 1971), that is, the frequencies of studies that had significant positive, significant negative, and insignificant results are determined and compared. Although this procedure is tempting and seems to deliver a clear finding, the sample sizes of the studies—besides the size of the effect another determinant of the probability of error—are not considered which can result in an inadequate conclusion regarding the general effect (but for a more elaborated version of vote counting see Bushman & Wang, 2009). Another approach for the aggregation of significance levels exists (Stouffer's z method; Stouffer, Suchman, Devinney, Star, & Williams, 1949) and has also been most widely criticized (but see Darlington & Hayes, 2000).

To characterize the third type of summarizing the findings of prior studies, Glass (1976) coined the term meta-analysis and defined it as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3). An advantageous characteristic of meta-analyses is that they are based on effect sizes that model the outcome of a given study without being (directly) influenced by the corresponding sample size. Detailed information on specialized effect size indices that can be applied to depict the findings of outcome evaluations is given in Manuscript #1 and Manuscript #2.

Two meta-analytical paradigms for the integration of individual effect sizes have been established and are utilized in different fields of psychology (and other disciplines). Whereas the *psychometric meta-analysis* (Hunter & Schmidt, 1990, 2004) has been developed and is often applied in industrial and organizational psychology, the procedure introduced by Larry Hedges and Ingram Olkin (1985) in their influential book *Statistical Methods for Meta-Analysis* is widely used in other fields, including evaluations of interventions, clinical psychology, and medicine. This approach suggests a weighted integration of the individual effects that is based on the so-called inverse variance weights (i.e., each effect is multiplied with the inverse of its squared standard error). Moreover, when random or mixed models (Hedges & Olkin, 1985; Hedges & Vevea, 1998; Raudenbush, 1994, 2009; Viechtbauer, 2004) are used, an additional component representing the random effects variance is incorporated in the inverse variance weight of each study. Further information on these models and the consequences of using them instead of fixed effects integrations is given in Manuscript #1 and Manuscript #2.

The meta-analytic process can be characterized as a six-step approach including the following elements:

1. Definition of criteria for the inclusion and exclusion of individual studies. After a researcher has decided to analyze a question meta-analytically, he has to specify detailed criteria (e.g., regarding suitable independent and dependent variables as well as concerning appropriate design features) that determine which studies are included in his analysis and which are not. It is important to communicate these criteria in the paper so that the reader knows about the exact nature of the meta-analyzed studies.

2. Literature search. After concrete rules for the inclusion and exclusion of studies have been stipulated, relevant literature (i.e., studies that satisfy all of the defined inclusion criteria) has to be found. In this regard, it is important to use a variety of different techniques (e.g., searches in disciplinary and multi-disciplinary databases as well as searches in the reference lists of the existing topic-related reviews and of the already identified primary studies) for the localization of research that is eligible for inclusion. This second step of a meta-analysis is of essential importance as the validity of the meta-analytic findings substantially depends on it. An unrepresentative sample (e.g., one that is mostly characterized by research that is easily accessible) could considerably bias the results and the corresponding conclusions.

3. Coding of the relevant studies. The information contained in the documents that were found and assessed as suitable has to be systematically coded. Two broad categories of coding variables are usually considered: study characteristics (e.g., type of publication, type of intervention, methodological characteristics) have to be specified and effect sizes have to be calculated by—typically—using the aggregate statistics provided in the documents.

4. Description of the research landscape. Since study features are coded, a systematic tabular overview of the existing studies on the topic of the meta-analysis can be given. Such a

description reveals the frequency distributions of essential characteristics (e.g., type of intervention, dependent variable(s), participants, and setting). It is a very precious "by-product" of a meta-analysis as it is based on exhaustive literature searches and therefore has the ability to depict research gaps and to stimulate new research.

5. Calculation of the estimated average true effect. After the studies were coded, the meta-analytic integration of the individual observed effects is conducted in order to get information about the average true effect, that is, the overall effect.

6. Moderator analyses. Finally, it is usually tested whether the magnitude of the effect sizes is influenced by certain study features. To be more precise, in the case of evaluations of social programs it is meta-analytically examined whether the impact of the investigated interventions on the dependent variable(s) of interest is moderated by further variables such as the specific type of intervention and methodological features.

Details concerning the procedures I used in the context of the respective steps are given in Manuscript #1 and Manuscript #2.

One final note: Originally, the term meta-analysis was introduced to describe the statistical portion of a research synthesis (i.e., the coding of effect sizes as well as step 5 and 6). However, in the course of time a more comprehensive understanding of the research instrument *meta-analysis* has been firmly established that also includes its preceding steps and implies that they are conducted in a highly structured and transparent fashion. Nonetheless, some authors as well as the Cochrane Collaboration and the Campbell Collaboration currently recommend to use the term *systematic review* to characterize reviews that are conducted in a structured way with regard to the first three steps and can either include an meta-analysis (systematic review with meta-analysis) or integrate the existing primary studies in a narrative way (systematic review without meta-analysis). Consequently, this interpretation opens up the possibility that a meta-analysis is conducted without a

systematic procedure in regard to step 1 to 3. However, since it is commonly accepted to use the term meta-analysis in a way that refers to the entire six-step process that I described, I prefer using it in accordance with that meaning.

1.5. Previous Reviews of the Effectiveness of Interventions to Reduce Ethnic Prejudice

Having briefly described the procedure of outcome evaluations and meta-analyses, I now come back to the actual topic under study. I have already mentioned that several reviews (e.g., Aboud & Levy, 2000; Bigler, 1999; Denson, 2009; Engberg, 2004; Gudykunst, 1977, 1979; Lynch, 1985, 1987; McGregor, 1989; 1993; McGregor & Ungerleider, 1993; Okoye-Johnson, 1999, Paluck & Green, 2009; Pedersen, Walker, Paradies, & Guerin, 2011; Pedersen, Walker & Wise, 2005, Pendry, Driscoll, & Field, 2007; Ponterotto & Pedersen, 1993; Rose, 1948; Schofield, 1995; Stephan, 1999; Stephan, Renfro, & Stephan, 2004; Stephan & Stephan, 2001; St. Jean, 2007; Williams, 1947) summarize the existing studies on the effectiveness of interventions to improve ethnic attitudes. As described in Manuscript #1 and Manuscript #2, the vast majority of these reviews is of narrative nature and—from a general point of view—can not provide a clear picture concerning the outcome of interventions to reduce ethnic prejudice. Although most of them conclude that interethnic attitudes can be improved, some are more pessimistic (see Manuscript #1 and Manuscript #2 for more information). In addition, no clear answer is given to the question of especially effective types of interventions.

Pettigrew and Tropp (2006, 2008; Tropp & Pettigrew, 2005a, 2005b) have conducted a series of meta-analyses in order to test whether contact is associated with reduced prejudice. The authors found clear evidence for their prediction. However, since these meta-analyses are mainly based on survey-research, they indicate that actual contact interventions can be expected to be effective but can not confirm this assumption.

A small number of meta-analyses (Denson, 2009; McGregor, 1989, 1993; McGregor & Ungerleider, 1993; Okoye-Johnson, 1999; Stephan, Renfro & Stephan, 2004) have investigated the impact of information interventions on ethnic attitudes. Supporting evidence was found, however, all of the existing meta-analyses are marked by severe shortcomings. For instance, they are respectively based on a very small number of included studies and only address specialized types of programs to reduce ethnic prejudice.

To summarize the current state of the research, it is unclear—although contact and information interventions can be seen as promising approaches—and has to be tested whether prejudice reduction programs really reach their aim. In addition, it has to be identified which types of interventions are the most beneficial.

1.6. The Present Work

The work that is described in this dissertation was conducted to test the effectiveness of contact and information interventions by the use of meta-analytic methods. Two meta-analyses were carried out.

The first meta-analysis is described in Manuscript #1. It aimed at testing the outcome of programs that are based on the intergroup contact theory and implemented in order to reduce ethnic prejudice by the use of face-to-face interaction between members of different ethnic groups or by initiating the experience of indirect contact. With this meta-analysis, we not only investigated the general impact of contact interventions but also tested further hypotheses. In this context, we predicted that: (1) contact programs are more effective for ethnic majorities than for ethnic minorities but have a positive impact for the latter, too; (2) contact interventions are also effective in regions that are marked by an intractable conflict; (3) ethnic prejudice is not only reduced by direct interethnic contact but also by variants of indirect contact; (4) the outcome of contact strategies is generalized in a sense that it is not

restricted to ethnic outgroup members who were met within the contact situation but also observable for the ethnic outgroup as a whole as well as for other ethnic outgroups.

The meta-analysis that is described in Manuscript #2 addresses the effectiveness of information interventions, that is, of programs that give some kind of input which is not based on intergroup contact. As already metioned, we introduce a typology for the classification of information initiatives. The taxonomy we advocate encompasses three axes: content, method, and duration. In addition, we specified and tested several hypotheses that directly relate to out typology. First, we expected that interventions that address empathy are more effective than programs that do not include such a content-component. Furthermore, we predicted that programs which (also) use active methods that involve the participants (i.e., discussion, small group activity, individual activity, or role play/simulation game) have a better outcome than programs that do not apply active but solely passive methods (i.e., reading, listening, receiving audio-visual information, or participating in lectures). Finally, we assumed that the length of information strategies is positively associated with their effectiveness.

The following sections contain the two manuscripts (Manuscript # 1: page 29– 137; Manuscript #2, page 138– 259). This dissertation concludes with a general discussion and a summary in German.

References

Aboud, F. E., & Levy, S. R. (2000). Interventions to reduce prejudice and discrimination in children and adolescents. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 269–294). Mahwah, NJ: Erlbaum.

Abrams, D., & Houston, D. M. (2006). *Equality, diversity and prejudice in Britain: Results from the 2005 National Survey*. Retrieved from http://kar.kent.ac.uk/4106/1/Abrams_KentEquality_Oct_2006.pdf

Ahmed, A. M., & Hammarstedt, M. (2008). Discrimination in the rental housing market: A field experiment on the Internet. *Journal of Urban Economics, 64*, 362–372.

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

Altemeyer, B. (1996). *The authoritarian spectre*. London, UK: Harvard University Press.

Bigler, R. S. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues, 55*, 687–705.

Brewer, M. B., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller & M. B. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 291–302). Orlando, FL: Academic.

Bushman, B. J. (2004). You've got mail: Using e-mail to examine the effect of prejudiced attitudes on discrimination against Arabs. *Journal of Experimental Social Psychology, 40*, 753–759.

Bushman, B. J., & Wang, M. C. (2009). Vote counting methods in meta-analysis. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 207–220). New York, NY: The Russell Sage Foundation.

Campbell, D.T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin, 54*, 297–312.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand McNally.

Campbell, D. T., & Stanley, J. C. (1966). E*xperimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Carlsson, M., & Rooth, D. O. (2007). Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics, 14*, 716–729.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Darlington, R. B., & Hayes, A. F. (2000). Combining independent p values: extensions of the Stouffer and binomial methods. *Psychological Methods, 5*, 496–515.

Denson, N. (2009). Do curricular and co-curricular diversity activities influence racial bias? A meta-analysis. *Review of Educational Research, 79*, 805–838.

Dessel, A., & Rogge, M. (2008). Evaluation of intergroup dialogue: A review of the empirical literature. *Conflict Resolution Quarterly, 26*, 199–238.

Dovidio, J. F., Gaertner, S. L., Nier, J. A., Kawakami, K., & Hodson, G. (2004). Contemporary racial bias: When good people do bad things. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 141–167). New York, NY: Guilford.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62–68.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt, Brace, Jovanovich.

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*, 473–524.

European Union Agency for Fundamental Rights (2010). *Annual report on fundamental rights*. Retrieved from http://fra.europa.eu/fraWebsite/research/publications/

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition. research: their meaning and use. *Annual Review of Psychology, 54*, 297–327.

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia, PA: Psychology Press.

Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher, 5*, 3–8.

Gudykunst, W. B. (1977). *The effects of an intercultural communication workshop on cross-cultural attitudes and interaction* (Unpublished doctoral dissertation). University of Minnesota.

Gudykunst, W. B. (1979). The effects of an intercultural communication workshop on cross-cultural attitudes and interaction. *Communication Education, 28*, 179–187.

Gurr, T. R. (1970). *Why men rebel*. Princeton, NJ: Princeton University Press.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93*, 388–395.

Hedges, L. V, & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V, & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.

Holmes, K., Murachver, T., & Bayard, D. (2001). Accent, appearance and ethnic stereotypes in New Zealand. *New Zealand Journal of Psychology, 30*, 79–86.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage Publications.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: correcting error and bias in research findings* (2nd ed.). Newbury Park, CA: Sage Publications.

Klink, A., & Wagner, U. (1999). Discrimination against ethnic minorities in Germany: Going back to the field. *Journal of Applied Social Psychology, 29*, 402–423.

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength Antecedents and consequences* (Vol. 4, pp. 1–24). Hillsdale, NJ: Erlbaum.

Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review, 41*, 429–471.

Lipsey, M. W., & Cordray, D. S. (2000) Evaluation methods for social intervention. *Annual Review of Psychology, 51*, 345–375.

Lynch, J. (1985). Human rights, racism and the multicultural curriculum. *Educational Review, 37*, 141–152.

Lynch, J. (1987). *Prejudice reduction and the schools*. New York, NY: Nichols.

Kaas, L., & Manger, C. (in press). Ethnic discrimination in Germany's labour market: A field experiment. *German Economic Review.*

Massey, D. S., & Lundy, G. (2001). Use of black english and racial discrimination urban housing markets: New methods and findings. *Urban Affairs Review, 36*, 452–469.

McGregor, J. (1989). *Strategies to reduce racial prejudice in students: A meta-analysis of research* (Unpublished doctoral dissertation). McGill University (Canada).

McGregor, J. (1993). Effectiveness of role playing and antiracist teaching in reducing student prejudice. *Journal of Educational Research, 86*, 215–226.

McGregor, J., & Ungerleider, C. (1993). Multicultural and racism awareness programs for teachers: A meta-analysis of the research. In K. A. McLeod (Ed.), *Multicultural*

education: The state of the art national study report #1 (pp. 59–63). Toronto, ON: University of Toronto, Faculty of Education.

Okoye-Johnson, O. N. (1999). *A meta-analysis of the effects of multicultural education on the racial attitudes of pre–k through grade 12 students: A comparison of curricular intervention and reinforcement dimensions in suburban and urban educational settings* (Unpublished doctoral dissertation). Old Dominion University.

Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: What works? A critical look at evidence from the field and the laboratory. *Annual Review of Psychology, 60*, 339–367.

Pedersen, A., Walker, I., Wise, M. (2005). Talk does not cook rice: Beyond anti-racism rhetoric to strategies for social action. *Australian Psychologist, 40*, 20–30.

Pedersen, A., Walker, I., Paradies, Y., & Guerin, B. (2011). How to cook rice: A review of ingredients for teaching anti-prejudice. *Australian Psychologist, 46*, 55–63.

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.

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

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

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*, 922–934.

Pettigrew, T. F., & Tropp, L .R. (2011). *When groups meet: The dynamics of intergroup contact*. London, UK: Routledge/Psychology Press.

Ponterotto, J. G., & Pedersen, P. B. (1993). *Preventing prejudice: A guide for counselors and educators*. Newbury Park, CA: Sage.

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology, 67*, 741–763.

Raudenbush, S. W. (1994). Analyzing effect sizes: Random effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 302–321). New York, NY: The Russell Sage Foundation.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models.In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–316). New York, NY: The Russell Sage Foundation.

Riach, P. A., & Rich, J. (2002). Field experiments of discrimination in the market place. *Economic Journal, 112*, F480–F518.

Rose, A. M. (1948). *Studies in reduction of prejudice: A memorandum summarizing research on modification of attitudes*. Chicago, IL: American Council on Race Relations.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59–82.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.

Schofield, J. W. (1995). Improving intergroup relations among students. In J. A. Banks, & C. A. M. Banks (Eds.), *Handbook on multicultural education* (pp. 628–634). New York, NY: Macmillan Publishing.

Schütz, H., & Six, B. (1996). How strong is the relationship between prejudice and discrimination? A meta-analytic answer. *International Journal of Intercultural Relations, 20*, 441–462.

Schweitzer, R., Perkoulidis, S. A., Krome, S. L., & Ludlow, C. N. (2005). Attitudes towards refugees: The dark side of prejudice in Australia. *Australian Journal of Psychology, 57*, 170–179.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi–experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Sherif, M. (1966). *Group conflict and cooperation*. London, UK: Routledge & Kegan Paul.

Sherif, M., & Sherif, C. W. (1969). *Social psychology*. New York, NY: Harper & Row.

Sidanius, J., & Pratto, F. (1999). *Social dominance*. Cambridge, UK: Cambridge University Press.

St. Jean, G. (2007). *Evaluation of intervention programs designed to address high school racial conflicts* (Unpublished doctoral dissertation). George Mason University.

Stephan, C. W., Renfro, L., & Stephan, W. G. (2004). The evaluation of multicultural education programs: Techniques and a meta-analysis. In W. G. Stephan & W. P. Vogt (Eds.), *Education programs for improving intergroup relations: Theory, research, and practice* (pp. 227–242). New York, NY: Teachers College Press.

Stephan, W. G. (1999). *Reducing prejudice and stereotyping in schools* (Multicultural Education Series). New York, NY: Teachers College Press.

Stephan, W. S., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–45). Mahwah, NJ: Erlbaum.

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

Stouffer, S. A., Suchman, E. A., Devinney, L. C., Star, S. A., & Williams, R. M. (1949). *The American Soldier*. Princeton, NJ: University Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Monterey, CA: Brooks/Cole.

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

Tropp, L. R., & Pettigrew, T. F. (2005a). Differential relationships between intergroup contact and affective and cognitive indicators of prejudice. *Personality and Social Psychology Bulletin, 31*, 1145–1158.

Tropp, L. R., & Pettigrew, T. F. (2005b). Relationships between intergroup contact and prejudice among minority and majority status groups. *Psychological Science, 16*, 951–957.

U.S. Department of Justice (2009). *Hate crime statistics of the Federal Bureau of Investigation*. Retrieved from http://www2.fbi.gov/ucr/hc2009/index.html

U.S. Department of State (2009). *The country reports on human rights practices.* Retrieved from http://www.state.gov/g/drl/rls/hrrpt/2009/index.htm

Viechtbauer, W. (2004). *Choosing between the fixed-, random-, and mixed-effects model in meta-analysis: An analysis of existing and new model selection methods* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Williams, R. M., Jr. (1947). *The reduction of intergroup tensions*. New York, NY: Social Science Research Council.

Zick, A., Pettigrew, T. F., & Wagner, U. (2008). Ethnic prejudice and discrimination in Europe. *Journal of Social Issues, 64*, 233–251.

Zick, A., Wolf, C., Küpper, B., Davidov, E., Schmidt, P., & Heitmeyer, W. (2008). The syndrome of group-focused enmity: The interrelation of prejudices tested with multiple cross-sectional and panel data. *Journal of Social Issues, 64*, 363–383.

2. Manuscript #1:

Do Direct and Indirect Contact Programs Improve Ethnic Attitudes? A Meta-Analytic Evaluation of Theory-Driven Interventions

Do Direct and Indirect Contact Programs Improve Ethnic Attitudes?

A Meta-Analytic Evaluation of Theory-Driven Interventions

Gunnar Lemmer and Ulrich Wagner

University of Marburg, Germany

Author Note

Gunnar Lemmer, Department of Psychology, University of Marburg; Ulrich Wagner, Department of Psychology and Center of Conflict Studies, University of Marburg.

This research was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft e.V., WA 1150/9-1/2) awarded to Ulrich Wagner.

Correspondence concerning this article should be addressed to Gunnar Lemmer, Department of Psychology, Section for Psychological Methods, University of Marburg, Gutenbergstr. 18, 35032 Marburg, Germany. Email: gunnar.lemmer@staff.uni-marburg.de.

Abstract

The present meta-analysis tests the effectiveness of interventions that are based on the intergroup contact theory and implemented to improve interethnic relations. Evaluations of direct (i.e., face-to-face) as well as indirect contact programs were included. The meta-analyzed contact-control comparisons model the effect of contact shortly after the interventions (k = 115, N = 10,591) or with a delay of at least one month (k = 23, N = 1,449). As hypothesized, the results indicate that theory-driven contact interventions improve ethnic attitudes. Also as predicted, the effect is larger for ethnic majorities, contact programs, however, have a positive impact on ethnic minorities as well. We demonstrate that contact interventions are also effective in the context of an intractable conflict. In addition, both direct and indirect contact programs have a positive outcome. Furthermore, not only attitudes toward individuals involved in the contact situation are improved but also toward the entire outgroup.

Keywords: meta-analysis, theory-driven intervention, intergroup contact theory, interethnic relations, ethnic attitudes

Do Direct and Indirect Contact Programs Improve Ethnic Attitudes?
A Meta-Analytic Evaluation of Theory-Driven Interventions

Recent studies demonstrate that negative relations between members of different ethnic groups are still a reality of social life (e.g., European Union Agency for Fundamental Rights, 2010; U.S. Department of Justice, 2009; U.S. Department of State, 2009) and reach the point of acts of hate crime. These facts indicate that knowledge of effective interventions that are able to improve interethnic relations in real-world settings outside the laboratory is badly needed.

In order to maximize the chances of success, interventions should be theory-driven. That is, they should be based on an empirically supported scientific theory that is able to explain the existence of negative intergroup relations or to predict how intergroup relations can be improved. Besides the criterion of being supported by empirical evidence, a further demand on an underlying theory is that its assumptions focus on modifiable variables and that they are convertible into real-world interventions.

Among social psychological theories of intergroup relations, the intergroup contact theory stands out as a particularly powerful one (Brown & Hewstone, 2005; Dovidio, Gaertner, & Kawakami, 2003; Pettigrew & Tropp, 2011). This approach is supported by an enormous body of research (e.g., Pettigrew & Tropp, 2006) and has recently been augmented by new theoretical aspects (e.g., Crisp, Stathi, Turner, & Husnu, 2009; Turner, Crisp & Lambert, 2007; Wright, Aron, & Brody, 2008; Wright, Aron, McLaughlin-Volpe, & Ropp, 1997). It directly provides a viable blueprint for action: In order to improve interethnic relations, persons with different ethnic backgrounds should be brought in direct (i.e., face-to-face) contact under structured conditions or should experience structured indirect (i.e., non-face-to-face) contact.

Given the supporting research on the underlying theory, planned interventions based on interethnic contact can be expected to have a positive impact on ethnic attitudes. However, their actual effectiveness in real-word settings has to be tested by means of outcome evaluations. Evaluations of programs that are based on the principles of intergroup contact are a very stringent test of the intergroup contact theory since they scrutinize the causal pathway from contact that is implemented under realistic conditions to improved ethnic attitudes. With regard to the explanatory power, a systematic quantitative review of evaluations is superior to a single evaluation study or the inspection of a few well-known evaluations. Unfortunately, up to now, a comprehensive quantitative research synthesis with focus on the effectiveness of contact interventions is not available.

The aim of this paper is to use meta-analytic techniques to examine the impact of contact programs on ethnic prejudice in a general way and, in doing so, to test the intergroup contact hypothesis and its conceptual extensions in a real-world interventional framework. To the best of our knowledge, this is the first quantitative review of actual interventions based on the intergroup contact theory.

Intergroup Contact Theory

Intergroup contact hypothesis. The basic principle of the intergroup contact hypothesis, "contact brings friendliness" (p. 12), was first mentioned by Robin Williams, Jr. (1947). Gordon Allport (1954) formally stated the intergroup contact hypothesis in his seminal book *The Nature of Prejudice*. He held that interactions between members of different groups can reduce prejudice. Allport (1954) limited the validity of his statement to situations with "equal status contact between majority and minority groups in the pursuit of common goals" (p. 281). He further specified that this "effect is greatly enhanced if this contact is sanctioned by institutional supports (i.e., by law, custom, or local atmosphere), and provided it is of a sort that leads to the perception of common interests and common

humanity between members of the two groups" (Allport, 1954, p. 281). Taken together, Allport (1954) characterized two necessary factors (equal status, pursuit of common goals) and two adjuvant aspects (sanction by institutional support, perception of common interests and common humanity) for contact to be effective.

Over more than half a century, an overwhelming body of research on intergroup contact has been developed. Pettigrew and Tropp (2006) summarized the existing literature quantitatively. Their meta-analysis included studies that were conducted before the year of 2001 and is primarily based on cross-sectional surveys that report correlations between the amount of self reported face-to-face contact with other groups and prejudice toward these groups. Altogether, Pettigrew and Tropp (2006) considered 515 studies with 713 samples and found that face-to-face contact is negatively correlated with prejudice (mean r = -.21 at the level of samples). Their synthesis also demonstrated that neither Allport's two necessary nor his two adjuvant conditions are essential for contact to be negatively associated with prejudice, but that the negative contact-prejudice relation is higher when they are satisfied.

Given the conclusive, primarily survey-based evidence for the contact hypothesis, it can be expected that interventions that are based on the principle of intergroup contact are effective. However, their actual impact in real-world settings has to be tested empirically.

Specific aspects of intergroup contact. Recently, two facets of the effect of intergroup contact have been given special attention. These aspects are of great importance to interventions and deal with the differential impact of contact on majority and minority members as well as with the effectiveness of contact in conflict zones.

Differences between ethnic majority and minority groups. An important question, which has been neglected for a long time, refers to the status relation between the ethnic groups involved in intergroup contact. More specifically, it can be asked whether the association between contact and interethnic relations is universally valid or whether it differs

for the ethnic majority (e.g., U.S. citizens with European-American roots) and the ethnic minority (e.g., U.S. citizens with African-American roots). Tropp and Pettigrew (2005) demonstrated in a further meta-analysis—again mainly based on survey-based research—that the negative association between face-to-face contact and prejudice is larger for ethnic majority samples (mean r = -.24) than for ethnic minority samples (mean r = -.18). A recent longitudinal study with data from Germany, Belgium, and England (Binder et al., 2009) replicated this finding in an even more drastic form. Although there is a promising approach to explain the mentioned discrepancy in effectiveness (see Pettigrew & Tropp, 2011; Tropp & Pettigrew, 2005), its cause has not yet been fully understood.

Based on the available empirical findings, it can be expected that the impact of contact programs is stronger for ethnic majority members compared to minority members, while they are effective for the latter group as well. However, up to now, this assumption has not yet been tested meta-analytically in the context of real-world interventions.

Contact in the context of intractable conflicts. Pettigrew and Tropp's meta-analyses do not consider whether a macro-level or societal conflict is involved. Salomon (2006) explicitly differentiates between initiatives that are implemented to improve interethnic relations in contexts of intractable conflicts and programs which are realized in settings where there are no such conflicts. Intractable conflicts are often also termed as protracted social conflicts that are specified by Azar (1990) as conflicts which are based on "ethnic hostilities crossed with developmental inequalities that have a long history and a bleak future" (p. 127). Rouhana and Bar-Tal (1998) characterize intractable conflicts as protracted, central in public life, violent, and perceived as irreconcilable. Examples of regions with a current intractable conflict or an intractable conflict in the recent past are the Middle East, South Africa, Northern Ireland, Cyprus, and Sri-Lanka. In such settings, negative interethnic relations not only adopt the shape of negative attitudes and discriminative tendencies but

often of severe acts of violence. Consequently, it can be assumed that contact interventions are "surrounded by a general atmosphere of hostility" in these contexts (Salomon, 2004, p. 262). This also holds true when the actual macro-level or societal conflict has already ended since it has governed everyday life for a long period of time, continually goes on in people's minds, and acts of violence and hostility are still observable (e.g., MacGinty, 2010; MacGinty, Muldoon & Ferguson, 2007). Thus, there is reason to believe that major obstacles (still) exist and that contact programs do not have a positive impact in such (former) conflict zones. This position is held by critics of the scope of intergroup contact (e.g., McGarry & O'Leary, 1995).

However, there is also good reason to assume that the principle of intergroup contact is universally valid and that contact interventions are effective even under the conditions of intractable conflicts (see also Wagner & Hewstone, in press). Survey-based research shows that contact is associated with reduced prejudice in (former) conflict regions like Northern Ireland (e.g., Paolini, Hewstone, Cairns, & Voci**,** 2004; Tam et al, 2008; Tausch, Hewstone, Kenworthy, Cairns, & Christ, 2007; Tausch, Tam, Hewstone, Kenworthy, & Cairns, 2007), South Africa (e.g., Dixon et al., 2010a; Dixon et al, 2010b; Durrheim, 2010; Gibson, 2006; Gibson & Claassen, 2010; Swart, Hewstone, Christ, & Voci, 2010; Tredoux & Finchilescu, 2010), and Sri Lanka (Malhotra & Liyanage, 2005). In addition, by means of narrative reviews, Salomon (2004, 2006, 2009) found support for the assumption that contact-based initiatives of peace building can have a positive impact in the Israeli-Palestinian context.

Taken together, it can be expected that contact programs improve ethnic attitudes even in areas with an intractable contact. Though, as described, there are substantial barriers and meta-analytic evidence does not yet exist.

Theoretical extensions of intergroup contact. Over time, the original contact hypothesis has been supplemented by several conceptual extensions with the result that the

contact approach can be regarded as an intergroup contact theory (Brown & Hewstone, 2005; Pettigrew, 1998; Pettigrew & Tropp, 2011). Among the aspects that have amended the intergroup contact hypothesis, some supplements are of special importance with regard to real-world interventions. They refer to indirect intergroup contact as well as to the generalization of contact effects.

Indirect intergroup contact. The term indirect intergroup contact describes the experience of intergroup contact that is not based on physical face-to-face interactions. We will outline three variants: extended, imagined, and virtual contact.

The extended intergroup contact hypothesis (Wright, Aron, & Brody, 2008; Wright, Aron, McLaughlin-Volpe, & Ropp, 1997) states that knowing that a member of the ingroup has a close relationship with a member of an outgroup results in improved attitudes toward the respective outgroup. At present, a number of laboratory experiments (e.g., Gómez & Huici, 2008; Kiu, 2006; Mazziotta, Mummendey, & Wright, 2011; Wright et al., 1997) as well as survey studies (e.g., Christ et al., 2010, Study 2; Turner, Hewstone, Voci, & Vonofakou, 2008) provide support for this hypothesis.

A further extension of the intergroup contact approach is the imagined intergroup contact hypothesis (Crisp, Stathi, Turner, & Husnu, 2009; Turner, Crisp, & Lambert, 2007). According to the authors, the mental simulation of a positive social interaction with a member of an outgroup leads to improved attitudes toward that group. Results from laboratory experiments (e.g., Crisp & Turner, 2009; Turner et al., 2007) support this postulate. There is also laboratory evidence demonstrating the effectiveness of imagined intergroup contact in the context of an intractable interethnic conflict (Husnu & Crisp, 2010).

Third, Amichai-Hamburger and McKenna (2006) point out the great potential of indirect contact between different ethnic groups by means of computer technology. The roots of this type of contact are located in the field of computer-mediated communication (CMT)

and information and communications technology (ICT), respectively. Here, we want to introduce the term *virtual contact* to characterize non-physical contact between real persons who are members of different groups via the use of computer technology. Studies supporting its positive impact already exist (e.g., Tavakoli, Hatami, & Thorngate, 2010).

Taken together, three variations of indirect contact have been conceptualized: extended, imagined, and virtual. The existing survey-based and laboratory evidence give reason to expect that real-world programs that are based on indirect contact are effective. However, the actual impact of these interventions has not yet been tested with a quantitative review.

A further conceptualization sometimes seen as an extension of intergroup contact originates from communication science and is labeled as the parasocial contact hypothesis (Schiappa, Gregg, & Hewes, 2005). The authors characterize their approach as a "communication analogue to Allport's (1954) contact hypothesis" (Schiappa et al, 2005, p. 92) and state that unidirectional contact between consumers on the one side and mass-mediated information from outgroup members on the other will improve attitudes of the receiver toward the involved outgroup. In contrast to extended, imagined, and virtual contact, we do not consider parasocial "contact" as contact in the sense of the intergroup contact theory but as an information-based approach. The reason for that is of conceptual nature: Parasocial contact is exclusively based on the transmission of information from a sender to a receiver without the possibility of mutual reactions. In contrast to direct and indirect contact, it does not formally include any type of bidirectional interethnic interaction, not even in the shape of imagined interactions.

Generalization of contact effects. The original intergroup contact hypothesis says nothing about if and how the impact of contact is generalized beyond the experienced situation and the participants involved in the contact. Regarding this, some critics (e.g.,

Rothbart, 1996; Rothbart & John, 1985, 1993) state that intergroup contact improves attitudes toward individual outgroup members who were met, but that it does not improve attitudes toward other members of that group or toward the entire outgroup. Consequently, the utility of the intergroup contact theory as a framework for interventions to improve interethnic relations would be severely limited.

Pettigrew (1998) differs, among other things, between the generalization to the entire outgroup and the generalization to other outgroups not involved in the contact situation. He (Pettigrew, 2009) termed the latter facet secondary transfer effect and distinguishes it from the primary transfer effect that characterizes the generalization to the target outgroup. Pettigrew (2009) as well as Pettigrew and Tropp (2011) provide substantial support for the existence of both effects in the context of face-to-face contact. In addition, a current study by Schmid, Hewstone, Küpper, Zick, & Wagner (in prep.) gives evidence for the secondary transfer effect with survey data from eight European countries. Recent studies also demonstrate that the secondary transfer effect occurs even in the context of indirect contact (Asbrock, Christ, Hewstone, Pettigrew, & Wagner, in prep; Harwood, Paolini, Joyce, Rubin, & Arroyo, 2011).

To summarize, there is evidence that contact effects are generalizable. Though, it has not yet been meta-analytically tested whether this applies to real-world contact interventions as well.

Theory-driven Interventions Based on the Intergroup Contact Theory

As a fundament of theory-driven interventions to improve interethnic relations, the intergroup contact theory has multiple beneficial properties. First, there is impressive empirical support for its assumptions that also demonstrates the generalization of contact effects. Second, it allows to directly derivate real-world interventions. The theory-based rationale is to bring people of different ethnic backgrounds in direct contact or to initiate

indirect contact, most suitable in a way that satisfies Allport's beneficial conditions. Third, due to the existence of the described extensions to different forms of indirect contact, the intergroup contact theory offers multiple options for interventions to realize (the experience of) interethnic interactions.

Having demonstrated the merits of the intergroup contact theory, we use the label *theory-driven interethnic contact intervention* to refer to carefully designed and implemented as well as highly structured programs in real-world settings that explicitly aim at the improvement of interethnic relations by the realization of direct or the experience of indirect interethnic interactions. In doing so, such interventions not merely establish structural opportunities for any kind of contact but warrant that direct and/or indirect interethnic contact really takes place and in the way it is intended to happen. Theory-driven contact interventions therefore typically satisfy Allport's beneficial conditions.

Given the enormous body of supporting survey-based and laboratory research on the association between contact and indices of intergroup relations, from the theoretical point of view contact interventions can be expected to be effective, that is, to improve interethnic relations. Their actual impact in applied settings, however, has to be verified by means of outcome evaluations. Contrary to an isolated consideration of some prominent outcome evaluations, a systematic and exhaustive synthesis of evaluations can verify the effectiveness of contact interventions in general and thus expand the external validity of the conclusions.

Existing contemporary reviews of contact programs (e.g., Dessel & Rogge, 2008; Sorensen, Nagda, Gurin, & Maxwell, 2009) are of qualitative nature and do not include interventions that are based on indirect contact. Meta-analytic reviews that quantitatively integrate the available research are superior to qualitative reviews which are vulnerable to subjective biases and cannot be used to analyze whether moderators exist, that is, whether the effectiveness of contact depends on the specific type of intervention and/or on the specific

context of the initiative. To the best of our knowledge, up to now, no meta-analysis on the topic exists that systematically and comprehensively integrates evaluations of actual interventions based on intergroup contact. The mentioned meta-analyses by Pettigrew and Tropp are primarily based on survey-based research and only include studies that were published before the year 2001. Moreover, they do not differentiate between contact in regions with an intractable conflict and regions without such a conflict and do not include studies on indirect contact. Additionally, older meta-analyses (Johnson & Johnson, 1989; Johnson, Johnson, & Maruyama, 1983, also reported in Johnson, Johnson, & Maruyama, 1984; Miller & Davidson-Podgorny, 1987) exist that, however, are not up-to-date and solely investigate the impact of one special type of face-to-face contact interventions, namely cooperative learning programs (see below), on interethnic relations. Furthermore, there are more recent meta-analyses (e.g., Roseth, Johnson, & Johnson, 2008) that were conducted to test the impact of cooperative learning programs on achievement or general peer relations without a focus on interethnic relations.

The Present Meta-Analysis

In light of the characterized circumstances, it is apparent that there is a need for an exhaustive and contemporary meta-analysis of the effectiveness of direct and indirect contact interventions. Such a systematic quantitative review should test if the assumed causal pathway from implemented contact to reduced prejudice in real-word settings generally exists. In addition, it should scrutinize the described new conceptual developments in the field of intergroup contact in an interventional framework.

The present work is the first meta-analysis that explicitly deals with these aspects. More specifically, it was conducted to test the following hypotheses.

Hypothesis 1: Contact interventions generally improve ethnic attitudes.

Previous research shows that contact is associated with reduced prejudice. We hypothesize that real-world programs that are based on the intergroup contact theory and implemented to improve interethnic relations have a positive impact.

In addition to the examination of the general outcome of contact interventions, we wanted to test more detailed hypotheses.

Hypothesis 2: Contact interventions are more effective for ethnic majorities. Nevertheless, they also have a positive impact on members of ethnic minorities.

Prior studies on face-to-face contact suggest that majority-minority contact has a stronger negative association with prejudice for majority members than for minority members. We expect to replicate this finding in the field of real-world interventions. More precisely, we assume that programs which bring ethnic majority and minority members together have a more positive impact on majority than on minority members. However, we predict contact interventions to be effective for ethnic minorities as well and therefore not being useless for them.

Hypothesis 3: Contact interventions not only reduce ethnic prejudice in "calm" settings but also in those that are marked by an intractable conflict.

As already described, there is reason to assume that contact programs have no impact in regions with (former) intractable conflicts (e.g., in the Middle East) as hatred and massive animosity toward the "other side" (still) exist. However, survey-based research and qualitative reviews give good reason to expect that contact interventions are not merely effective in calm regions but even in areas with an intractable conflict (in the recent past). We thus hypothesize programs based on the intergroup contact theory to have a positive impact in either setting.

Hypothesis 4: Both direct and indirect contact interventions improve ethnic attitudes.

Based on the underlying theoretical framework and the presented research on it, we expect that the positive effect of contact interventions is not restricted to direct, face-to-face programs but can also be verified regarding indirect contact experiences.

Hypothesis 5: The effect of contact interventions generalizes to the entire target outgroup as well as to outgroups not involved in the interventions.

Critics of the use of intergroup contact argue that contact merely improves relations between individuals who have met in the contact situation. Though, convincing empirical evidence for generalized contact effects exists. Therefore, we hypothesize that the impact of contact programs is not restricted to personal relationships with individual outgroup members but generalizes to attitudes toward the entire target outgroup (primary transfer effect) and toward outgroups which are not represented in the contact program (secondary transfer effect).

Taken together, with the present meta-analysis we intend to expand the current state of knowledge both on the intergroup contact theory itself and on interventions to improve interethnic relations by: (a) testing the causal effect of structured contact in real-world settings on ethnic attitudes; (b) investigating whether the majority-minority difference in the impact of contact can be replicated in the context of applied contact interventions while contact being also effective for ethnic minority members; (c) verifying the expected outcome of contact interventions in regions with an intractable conflict; (d) testing whether supportive evidence for the effectiveness of indirect contact interventions exists; (e) examining whether the effect of real-world contact programs generalizes to the entire outgroup whose members have been involved in the contact situation as well as to outgroups not involved. To the best of our knowledge, the present meta-analysis is the first to test the described hypotheses.

Method

Inclusion Criteria and Exclusion of Studies

We specified the population of studies that are eligible for inclusion in the present meta-analysis by means of several criteria. These inclusion criteria refer to five domains: independent variable, dependent variable, evaluation design, available data, and language.

1. Independent variable. To be eligible for consideration, a study has to evaluate the effectiveness of a theory-driven interethnic contact intervention. These interventions are characterized here as planned programs that: (a) are carefully designed, structured, and implemented in real-world settings outside the laboratory; (b) have the explicit objective to improve interethnic relations by the establishment of direct or indirect interethnic contact; (c) do not just implement (macro-structural) opportunities for any kind of contact by inducing physical proximity but warrant a controlled contact situation, that is, interethnic contact has to take place in the intended shape for the vast majority of the participants (i.e., for at least 75% of the participants).

These requirements rule out, for instance, the inclusion of macro-level programs (e.g., desegregation initiatives and housing projects) that open up opportunities for any kind of contact but do not satisfy the third criteria (e.g., Deutsch & Collins, 1951; Horowitz, 1936; Singer, 1966, 1967; Spangenberg & Nel, 1983; Wilner, Walkley, & Cook, 1955). For the same reason, we did not include student exchange and tourism studies (e.g., Amir & Ben-Ari, 1985; Anastasopoulos, 1992; Merkwan & Smith, 1999; Patterson, 2006; Pizam, Fleischer, & Mansfeld, 2002), interethnic roommate studies (e.g., Shook & Fazio, 2008; Van Laar, Levin, & Sidanius, 2008; Van Laar, Levin, Sinclair, & Sidanius, 2005), and evaluations of summer camps with ethnically mixed participants but without structured interethnic encounters (e.g., Acuff, 1975; Amir & Garti, 1977; D'Agostino, 1980; Mann, 1959, 1960; Mussen, 1950). Moreover, we excluded studies that intended to establish interethnic small groups (with six or

less members) but cannot warrant structured interethnic contact for at least 75% of the participants because, for example, more than 90% of the participants were members of the ethnic majority (e.g., Mingleton, 1993; Santos Rego & Moledo, 2005). Furthermore, we ruled out the inclusion of studies that were not realized in real-world contexts but in laboratory settings (e.g., Ashmore, 1969; Ioannou**,** 2009; Landis, Brislin, & Hulgus, 1985) as well as of studies that examine contact interventions that do not have an interethnic context but are, for instance, focused on relations between disabled and non-disabled persons or on relations between persons of different generations or differing sexual orientations (e.g., Cameron & Rutland, 2006; Hannon, 2004; Krahé & Altwasser, 2006). Besides, evaluations that merely report combined results of interethnic and non-interethnic contact programs (e.g., Nagda, Gurin, Sorensen, Gurin-Sands, & Osuna, 2009) and studies that are solely survey-based without any planned implementation of contact (e.g., Berryman-Fink, 2006; Liebkind, Haaramo, & Jasinskaja-Lahti, 2000; Sabar, Yogev, & Alper, 1987) were excluded.

2. Dependent variable. Relevant studies have to evaluate the impact of a contact intervention with at least one indicator of ethnic prejudice. Appropriate indicators focus on the affective (e.g., liking), cognitive (e.g., beliefs), or on the conative (e.g., associations with intended behavior) dimension of ethnic attitudes. We also decided to include studies that concentrate on other facets of interethnic relations such as actual behavior and implicit attitudes (i.e., attitudes that are typically unacknowledged or outside of awareness) toward ethnic outgroups.

When the results are reported separately for different subscales or items, we only regarded interethnic subscales or items and disregarded subscales or items that do not focus on interethnic relations. In case an instrument contains both interethnic items and items without interethnic reference and results are just reported for the overall scale, we decided to include it when at least 75% of the items have a clear reference to interethnic relations. This

criterion is not satisfied by many instruments whose titles actually suggest that they—at least primarily—measure interethnic relations. Examples are the "MAKKS–Multicultural Counseling Awareness, Knowledge and Skills Survey" (D'Andrea, Daniels, & Heck, 1991) or the "CCAI–Cross-Cultural Adaptability Inventory" (Kelly & Meyers, 1995, 1999).

We excluded evaluations that did not utilize at least one measure with a clear interethnic basis but exclusively applied dependent variables such as: attitudes toward the own ethnic group and racial/ethnic identity (e.g., Foerster, 1981; Hegarty, 1975; Reddick-Gibson, 1999), centrality of the concept of race (e.g., Nagda, Kim, & Truelove, 2004; Nagda & Zúñiga, 2003), tolerance for general diversity (e.g., Rouse, 2001; Seguin, 2002), dogmatism (e.g., Allgood, 1998; Carl & Jones, 1972), authoritarianism (e.g., Katz & Benjamin, 1960; Webster, 1958), general peer relations (e.g., Blaney, Stephan, Rosenfield, Aronson, & Sikes, 1977; Moskowitz, Malvin, Schaeffer, & Schaps, 1983, 1985; Slavin, 1978; Stevens & Slavin, 1992), as well as attitudes toward language learning and proficiency in foreign languages (e.g., Genesee, 1978; Kuhlemeier, van den Bergh, & Melse, 1996). In addition, studies were not considered relevant when they merely include measures of self-assessed multicultural skills and knowledge like, for instance, the "MCAS–Multicultural Counseling Awareness Scale" (Ponterotto et al, 1996). Within the scope of evaluations of interventions, such self-estimates of skills and knowledge with items like "I am aware of culture-specific, that is culturally indigenous, models of counseling for various racial/ethnic groups" are prone to many biases, for example, due to a potential motivation to reduce cognitive dissonance after investing in the participation in an intervention. Finally, in contrast to previous meta-analyses of cooperative learning techniques, we excluded studies that contain as the only interethnic dependent variable indices of interethnic interactions during the time period of the intervention (e.g., Katz, Goldston, & Benjamin, 1958; Shachar &

Sharan, 1994). We view such indices rather as a manipulation check since interethnic on-task behavior is a substantial element of the intervention itself (see below).

3. Evaluation design. We accepted studies with the following evaluation designs: *randomized posttest only with control (POWC)*, *pretest-posttest with control (PPWC)*, and *pretest-posttest single group (PPSG)*. The randomized POWC-design includes both an intervention group, whose members participate in the intervention, and a control group. The two groups are only tested after the end of the intervention. When randomization of participants to conditions is possible, a pretest is not necessarily needed and a POWC-design has no severe disadvantages concerning internal validity. However, in case an individual-based randomization is not feasible, potential differences between the groups can typically not be adequately controlled[1] which results in serious threats to internal validity (Shadish, Cook, & Campbell, 2002). Therefore, we did not consider studies that utilized a POWC-design and did not randomly assign individuals to conditions. In addition to evaluations with a randomized POWC-design, we included studies with a randomized or non-randomized PPWC-design. This design also involves an intervention and a control group. Both groups are measured at least at two points of time: before the program has started and at least at one point of time after the intervention has ended. Since the PPWC-design opens up the opportunity to control for pretest differences, it is especially beneficial when randomization of participants to the treatment or control group is not possible. Finally, we accepted evaluations with a PPSG-design. This design does not include a control group but just an intervention group that completes a pretest and at least one posttest after the program has ended. The PPSG-design is susceptible to substantial threats to internal validity as the results obtained with this design could not only reflect the intervention effect but also, for instance, a possible time effect (maturation and/or history, see Shadish, Cook, & Campbell, 2002). Nonetheless, we decided to include evaluations using this design. Intervention-based research

often cannot satisfy the standards of laboratory research and is typically subject to a number of severe restrictions such as of monetary and logistical nature. In addition, withholding an intervention can be ethically problematic. Hence, there is often no alternative to the application of a PPSG-design. To not include studies with this methodological weaker design from the beginning would mean to not consider a great deal of the existing evaluation research on the topic and to discard evaluations of many carefully constructed contact interventions (e.g., Bar-Natan, Rosen, & Salomon, 2010; Connolly, 1992). However, evaluations using a randomized POWC-design or a PPWC-Design are analyzed separately from studies using a PPSG-design (see below).

We excluded studies that are solely of qualitative nature without giving an opportunity to quantify their results (e.g., Franzen, 2009; Risberg, 1972; Wright & Tolan, 2009). In addition, we did not include one-shot case studies that only investigated subjective impressions of possible intervention-induced changes (e.g., Spearmon, 1999). Besides, we excluded evaluations that used ex post facto designs in which there was no control over the contact situation and inferences were made without manipulating the independent variable under study (e.g., Alderfer, Alderfer, Bell, & Jones, 1992; Maoz & Ellis, 2008; Muthuswamy, Levine, & Gazel, 2006; Spanierman, Neville, Liao, Hammer, & Wang, 2008; Wright & Tropp, 2005). Moreover, we did not consider studies that do not contain a pretest and compare different interventions at the time of the posttest without having an untreated control group (e.g., Cohen, Lockheed, & Lohman, 1976; Cohen & Roper, 1972; Johnson, Johnson, Tiffany, & Zaidman, 1984; Rzoska & Ward, 1991; Tam, 2001) or that compare pretest and posttest values of different samples of participants (e.g., Clore, Bray, Itkin, & Murphy, 1978; Eaton & Clore, 1975).

4. Available data. We decided to only include studies that allow calculating an effect size with a sufficient degree of precision. Determination of effect sizes for interventions is

ideally based on means, standard deviations, and sample sizes of the respective groups at the respective points of measurement (see equation 1 to 3 below). Unfortunately, these basic statistics are in many cases not provided in research documents. However, effect sizes can be also calculated from other statistics (e.g., t-test statistics) by using transformation formulas (e.g., Lipsey & Wilson, 2001). Nonetheless, an accurate determination of effect sizes was sometimes not possible with the information given in the papers. In these cases, the author(s) were contacted—if contact data were available—and we asked for additional information.

We excluded evaluations when statistical information to calculate an effect size was not available but, for instance, only results of complex multifactorial ANOVA's or MANOVA's with repeated measurement (e.g., Desai-Patel, 2005; Geffner, 1978; Kelly, 1971a, 1971b). In this regard, even studies that are often cited do not allow an appropriate determination of effect sizes that adequately mirror their results (e.g., Cook, 1971, 1984, 1985; Liebkind & McAllister, 1999; Walker & Crogan, 1998). In addition, we excluded studies with a PPSG-design that only provide means and standard deviations for samples whose sizes vary between the pre- and the posttest (e.g., Doubilet, 2007, Study 1; McAlister, Ama, Barroso, Peters, & Kelder, 2000). Moreover, we decided not to include evaluations (e.g., DeVries, Edwards, & Slavin, 1978) that only report their findings in form of proportions of interethnic (friendship) nominations—that is, the total number of interethnic choices devided by the total number of choices—so that the level of analysis is not the single participant but refers to the aggregated choices of all participants. We did this in contrast to prior meta-analyses of cooperating learning activities since the calculation of an appropriate effect size indicator is not possible for these studies.

5. Language. Finally, an additional formal criterion was set. We restricted the population of studies that are eligible for inclusion in the present meta-analysis to those that are reported in English or German.

To comprehensively capture the existing evaluation research on the topic, we did not specify criteria regarding the age of the participants, the point of time the study was conducted, and the country in which the intervention was implemented.

Search for Relevant Literature

In order to get as close as possible to the identification of the entire population of existing studies that satisfy the described inclusion criteria, we looked for relevant documents with five different strategies: searches in databases, manual searches in topic-related journals, consultation of organizations and experts, searches in conference proceedings, and inspection of reference lists. The searches were performed repeatedly. The final search took place in January 2011 to systematically cover documents that were written up to the year 2010.

1. Searches in databases. First, we looked for relevant evaluations in international databases of multiple scientific disciplines: Psychology (e.g., PsycINFO, PSYNDEXplus), Education (e.g., ERIC - Education Resources Information Center, ERS - Education Research Complete, FIS - Bildung Literaturdatenbank), Social Sciences (e.g., Sociological Abstracts, SSCI - Social Sciences Citation Index, Social Services Abstracts), Media- and Communication Science (Communication & Mass Media Complete), Sports Science (SPORTDiscus), and Medicine (PubMed). Second, we queried broad multidisciplinary international databases (e.g., Google Scholar, Scirus, WorldCat). Third, in order to find as many unpublished documents as possible, we searched in multidisciplinary databases for dissertations and master's theses. Regarding this, we looked for documents in international databases (e.g., NLTD - Networked Digital Library of Theses and Dissertations) and a multiplicity of country-specific databases, for example, ProQuest Dissertations & Theses (USA), Theses Canada, ULI - Israel Union Catalog, Australisian Digital Theses Program, Index to Theses (UK), DATAD - Database of African Theses and Dissertations, DissOnline (Germany), DiVA - Academic Archive On-line (Scandinavia), and NARCIS - National

Academic Research and Collaborations Information System (the Netherlands). Fourth, specialized databases containing grey literature were searched (e.g., OpenSIGLE, NTIS - National Technical Information Service, NCJRS Abstracts Database, DTIC Public technical reports, CPCI - SSH - Conference Proceedings Citation Index, Social Sciences & Humanities).

Within the databases, whenever possible, we utilized multiple structured and complex search algorithms in order to find studies that are eligible for inclusion. The algorithms we used were composed of numerous target words which refer to different components[2] of relevant studies. The arrangement of the created search algorithms can be illustrated with the following example: (a) contact component (e.g., contact*, cooperat*, interact*, dialogu*, encount*), (b) intervention component (e.g., intervent*, treatment*, training*, workshop*, program*, curriculum*, camp*), (c) ethnic component (e.g., ethnic*, racial*, cultural*, internat*), (d) dependent variable component (e.g., prejudic*, attitud*, stereotyp*, anxiet*, discriminat*, aggress*, violen*, relation*), and (e) evaluation component (e.g., evaluat*, effect*, outcome*, impact*, result*, reduc*, foster*, improv*). Within the five blocks we used a multiplicity of synonyms because, for instance, the words “contact”, “evaluation”, and “intervention” are often not mentioned in the titles and abstracts of relevant documents. Within the components, the synonyms were combined with the boolean operator *or*, the blocks were linked with the boolean operator *and*, so that entries in the databases were obtained that contain in each of the five components at least one of the synonyms.

2. Manual searches in topic-related journals. As an accompanying measure, journals that are related to the topic of study (e.g., Group Processes and Intergroup Relations, Intercultural Education, International Journal of Intercultural Relations, Journal of Applied Social Psychology, Journal of Educational Psychology, Journal of Ethnic and Migration Studies, Journal of Peace Education, Journal of Social Issues, Peace and Change, Peace and

Conflict, Race and Social Problems, Small Group Research) were manually searched. We included this component in order to detect further potential relevant literature that, for some reason, could not be detected with our search algorithms for the databases.

3. Consultation[3] of organizations and experts. A call for papers was repeatedly sent via the listservs of a multiplicity of topic-related scientific organizations from multiple disciplines (e.g., Divisions of the American Psychological Association, Divisions of the American Educational Research Association, Divisions of the German Psychological Society, European Evaluation Association, National Communication Association). Furthermore, we individually contacted scientific experts in the field of study from many countries. Third, we wrote to practitioners, evaluators, and organizations all over the world that are directly connected with interventions to improve interethnic relations (e.g., Building Bridges for Peace, Cyprus Youth Council for International Cooperation, Israel/Palestinian Center for Research and Information, Seeds of Peace).

4. Searches in conference proceedings. We also looked for relevant research in the proceedings of topic-related conferences (e.g., ICPRI Peace education Conferences, ISPP annual meetings).

5. Inspection of reference lists. In addition to the other approaches, we systematically searched the bibliographies of previous reviews that are—more or less—related to the topic under study (e.g., Amir, 1969, 1976; Black and Mendenhall, 1990; Bigler, 1999; Denson, 2009; Dessel & Rogge, 2008; Ford, 1986; Gudykunst, 1977, 1979; Johnson & Johnson, 1989; Johnson, Johnson, & Maruyama, 1983, 1984; McGregor, 1989; 1993; Miller & Davidson-Podgorny, 1987, Okoye-Johnson, 1999; Paluck & Green, 2009; Pettigrew & Tropp, 2006; Schofield, 1995; Sorensen, Nagda, Gurin, & Maxwell, 2009; Stephan, 1999; Stephan, Renfro, & Stephan, 2004; Stephan & Stephan, 2001; St. Jean, 2007; St. John, 1975; Williams, 1947)

as well as the reference list of each potentially relevant document that we found via the other search strategies.

Coding Procedure

The documents that were assessed as relevant were subjected to a detailed coding procedure consisting of two components: study characteristics and effect sizes.

Study characteristics. We coded formal characteristics, characteristics of the intervention, as well as characteristics of the methodological quality and the dependent variables. A list of the most central study characteristics can be seen in Table 1.

1. Formal characteristics. We registered the year the document was written, the publication status of the document (journal article, book/book chapter, dissertation/master thesis, unpublished evaluation report), the country of the first author, the number of pages, and whether further documents[4] exist that are directly related to the respective document.

2. Characteristics of the intervention. Here, a central variable is the type of intervention. We differentiated between two types of direct (contact meetings, cooperative learning programs) and three types of indirect (extended, imagined, virtual) interethnic contact interventions.

We term the first type of direct contact programs *contact meetings* to characterize a group of interventions that bring persons with different ethnic roots together in order to explicitly address the relations between the involved ethnicities. This is typically done by initiating structured intergroup discussions and dialogues. As a second type of direct contact interventions, *cooperative learning programs* instruct persons with a different ethnic background to work together cooperatively in small groups on a common learning aim or work product that does not relate to interethnic relations. In the context of cooperative learning initiatives, interethnic relations and group memberships are typically not highlighted. In this regard, cooperative learning programs systematically differ from contact meetings that

explicitly deal with the relationships between the involved groups and make group memberships salient.

In accordance with the described approaches of indirect contact, we classify indirect contact programs into three sub-types: extended, imagined, and virtual contact interventions. *Extended contact interventions* provide picture or written stories, radio plays, or films that explicitly display friendships or positive relations between at least one member of the own ethnic group and at least one member of an ethnic outgroup. *Imagined contact interventions* require participants to mentally simulate positive interactions with members of other ethnic groups. The programs we label *virtual contact interventions* initiate systematic exchange and discussion activities between members of different ethnic groups by means of computer technology.

Additionally, we classified in which country the program was implemented, how long it lasted (e.g., in net intervention hours, calculated as hours per intervention day x numbers of intervention days), and whether it was implemented[5] by virtue of research interests or solely for practical reasons to solve social problems. Besides the already reported variables, we registered the age and sex (in % female) of the participants as well as their ethnic backgrounds. Concerning the status relation between the ethnic groups involved, we classified whether the sample of participants the effect size is based upon had a majority position (e.g., European Americans, when involved in a program together with African Americans), had a minority position (e.g., African Americans, when involved in an intervention together with European Americans), represents a mixture of a majority and a minority position (e.g., when European Americans' and African Americans' data are reported in an aggregate form without differentiating between the two groups), or whether no direct status relation can be inferred (e.g., Americans, when involved in a contact intervention together with Europeans). Moreover, we registered whether the intervention was realized in

the context of a (former) intractable conflict (e.g., in the Middle East, South Africa, Cyprus) or whether this was not the case (e.g., when European Americans were brought in contact with African Americans or when Americans participated in a contact intervention together with Europeans).

3. Characteristics of the methodological quality and the dependent variables. With regard to the methodological quality, we, for example, coded the research design (POWC, PPWC, PPSG), the level (individual, group) and the type (e.g., randomized) of assignment of participants to conditions, the type of the control group (placebo treatment without interethnic reference, no treatment), the sample sizes, the extent of attrition in the intervention as well as in the control group (i.e., % loss from the initial sample to the sample that is considered in the data), whether a follow-up test took place, the time intervals between the end of the intervention, the posttest and the follow-up test, as well as the total number of items that were used to measure interethnic relations (as a proxy for the reliability).

The dependent variables were classified on three dimensions: content (i.e., cognitive, affective/behavioral, or mixed), type of measure (e.g., Likert attitude items, semantic differentials), and level of generalization. In reference to content, the categorization was restricted to the mentioned indices of explicit attitudes since no study focusing on actual, not self-reported behavior toward ethnic out-groups or on implicit ethnic attitudes was found. In addition, affective and conative variables were assigned to the same category as they could often not be adequately separated. With regard to the level of generalization, we classified whether the variable is targeted on individuals that were involved in the intervention or control group, is directed toward the entire outgroup that was represented by individuals in the intervention, toward ethnic outgroups in general without a specification (“other ethnic groups”), toward specific ethnic outgroups that were not represented in the intervention, or whether the variable represents a mixture of different generalization levels.

The coding process was based upon a detailed coding manual with a specification of the respective variables and categories, explicit coding rules, and typical examples. In order to evaluate the quality of the developed coding system, we specified the interrater reliability. For this, we trained a further coder who coded a random sample of 20% of the included intervention-control comparisons. The agreement was calculated by using Cohen's κ (Cohen, 1960). This index corrects the agreement for the accordance expected by chance. The results showed that Cohen's κ for the individual study characteristics ranges from 0.80 (for reason of implementation, i.e., scientific interests vs. practical reasons) to 1.0 (e.g., for intractable conflict vs. no intractable conflict and direct vs. indirect contact). The interrater reliability, therefore, can be classified as "substantial agreement" to "almost perfect agreement" (Landis & Koch, 1977).

Effect sizes. We have already noted that three evaluation designs were accepted: randomized posttest only with control (POWC), pretest-posttest with control (PPWC), and pretest-posttest single group (PPSG). Therefore, we decided to use three effect size indices to display the results of the included studies: Hedges's *g*, Morris's *g*, and Becker's *g*. These indices are sophisticated modifications of the standardized mean difference, also known as Cohen's *d*.

1. Hedges's *g* (Hedges, 1981, 1982; Hedges & Olkin, 1985). We applied Hedges's *g* to model the findings from evaluations with a randomized POWC-design. This index is given by

$$g_i^{Hedges} = \left(\frac{(\bar{Y}_i^{post,I} - \bar{Y}_i^{post,C})}{\sqrt{\frac{(n_i^{post,I}-1)(s_i^{post,I})^2 + (n_i^{post,C}-1)(s_i^{post,I})^2}{n_i^{post,I} + n_i^{post,C} - 2}}} \right) \left(1 - \frac{3}{4(n_i^{post,I} + n_i^{post,C}) - 9} \right). \quad (1)$$

Hedges's *g* for the *i*th study initially subtracts the posttest mean (denoted by $\bar{Y}$) of the control group (denoted by *C*) from the posttest mean of the intervention group (denoted by *I*).

The resulting term is divided by the pooled posttest standard deviation of both groups. This index is identical with Cohen's *d*. Hedges's *g* yet multiplicatively adds a correction factor. This factor prevents effect sizes from studies with small samples from being upwardly biased and approximates "1" with increasing samples sizes.

2. Morris's *g* (Morris, 2008). By the use of a simulation study, Morris (2008) verified the quality of several effect indices that are able to model the complete information that is provided by studies with a PPWC-design. It became evident that one index has especially beneficial properties, it is defined as

$$g_i^{Morris} = \left(\frac{(\bar{Y}_i^{post,I} - \bar{Y}_i^{pre,I}) - (\bar{Y}_i^{post,C} - \bar{Y}_i^{pre,C})}{\sqrt{\frac{(n_i^{pre,I} - 1)(s_i^{pre,I})^2 + (n_i^{pre,C} - 1)(s_i^{pre,C})^2}{n_i^{pre,I} + n_i^{pre,C} - 2}}} \right) \left(1 - \frac{3}{4(n_i^{pre,I} + n_i^{pre,C} - 2) - 1} \right) \quad (2)$$

and termed here as *Morris's g*. For the *i*th study, it initially subtracts the pretest mean from the posttest mean (respectively denoted by $\bar{Y}$) both within the intervention group (denoted by *I*) and within the control group (denoted by *C*). The difference within the control group is then subtracted from the difference within the intervention group. The resulting term is divided by the pooled standard deviation at the time of the pretest and multiplied by a correction factor. Morris's *g* can be applied exclusively to display the findings from studies with a PPWC-design. For this design, it has—compared to Hedges's *g* and Becker's *g*—the advantage that it models the intervention effect and simultaneously controls for pretest differences between the intervention and control group. Incorporating pretest results in an effect size for evaluation studies can be seen as the "standard of accuracy" (Carlson & Schmidt, 1999, p. 852).

3. Becker's *g* (Becker, 1988). Regarding studies with a PPSG-design, we used Becker's *g*. It is given by

$$g_i^{Becker} = \left(\frac{\bar{Y}_i^{post,I} - \bar{Y}_i^{pre,I}}{s_i^{pre,I}} \right) \left(1 - \frac{3}{4(n_i^{pre,I} - 1) - 1} \right). \quad (3)$$

For the *i*th study, this index subtracts the pretest mean (denoted by $\bar{Y}$) of the intervention group (denoted by *I*) from the posttest mean of that group and divides the resulting term by the pretest standard deviation (denoted by *s*) of the intervention group. Analogue to Hedges's *g* and Morris's *g*, a correction factor is multiplicatively added.

Each of the three[7] indices represents the difference between an intervention and a control condition in standard deviation units. Positive values indicate a positive contact effect[8].

Meta-Analytic Methods

Preliminary processing of effect sizes.

Clustering of effect sizes. We determined to meta-analytically integrate the three effect size indices in two separate clusters: a *primary* and a *secondary cluster*. The results of studies with a randomized POWC-design or a PPWC-design are assumed to represent the intervention effect in a way that is not influenced by possible pretest differences between the intervention and the control group as well as not influenced by possible time effects[9]. According to this, we infer that Hedges's *g* and Morris's *g* estimate the intervention population effect without being biased by pretest differences and time effects. Therefore, justified by research on meta-analytic methods (Morris & DeShon, 2002), we analyzed the two designs and effect size indices, respectively, together in a primary cluster. On the contrary to studies with a randomized POWC-design or a PPWC-design, the results of evaluations with a PPSG-design are assumed to reflect both the intervention effect and a time effect. We thus consider the findings of these studies and Becker's *g*, respectively, as biased estimators of the intervention effect. Nevertheless, as previously mentioned, they constitute a

substantial body of research on the effectiveness of real-world contact interventions. Hence, we analyzed them in a separate secondary cluster.

Unit of analysis. Whenever possible, we calculated separate effect sizes for different ethnic groups and age groups within the included studies. Determining detailed effects for different ethnic groups within the studies (e.g., for European Americans as well as for African Americans) enabled us to optimally test a majority-minority divergence as predicted in hypothesis 2. Computing separate effects for different age groups within the studies allowed a better test of a possible association of intervention effects with the participant's age. Since a given study could contribute with more than one effect size—each being based on a different sample of persons—to the meta-analytic integrations, we did not use study but *intervention-control comparison* as the unit of our meta-analytic tests. This term refers to the contrast between an intervention group and a control condition, with the latter being an actual control group or a control baseline condition in a study with a PPSG-design.

Elimination of stochastic dependencies. Prior to a meta-analytic integration, stochastic dependencies between the individual effect sizes have to be eliminated (Hedges & Olkin, 1985; Matt & Cook, 2009). This means that a given sample of participants must contribute with just one effect size estimate to a given meta-analytic integration. In order to create a distribution of stochastic independent data points within the primary and within the secondary cluster, we used several procedures.

1. When a document reports separate results for multiple relevant dependent variables for the same sample of participants, we calculated an effect size for each relevant measure and then aggregated[10] the individual effect sizes within the sample (i.e., within the intervention-control comparison) prior to the meta-analytic integration to obtain a single effect size. This procedure was also utilized in many other meta-analyses (e.g., Albarracín et al., 2005; Durantini, Albarracín, Mitchell, Earl, & Gillette, 2006; Mitte, 2005; Wilson &

Lipsey, 2007; Wilson, Lipsey, & Derzon 2003). Although information about specific features of the individual dependent variables are lost, we preferred this method to others (e.g., randomly selecting one effect size per sample) since the aggregate of relevant measures of ethnic prejudice mirrors the effectiveness of the respective intervention the best.

2. Some documents contain findings regarding different interventions and just one control group (e.g., Cameron, Rutland, Brown, & Douch, 2006). Incorporating multiple effect sizes in these cases—a separate effect size for the comparison of each intervention group with the same control group—would partially violate the assumption of independence[11] (Becker, 2000; Matt & Cook, 2009). To avoid biased estimates, we decided to systematically select[12] the contact condition for an intervention-control comparison that—according to the hypothesis of the author(s)—was a priori expected to be the most effective and therefore can be assumed to be the "optimal" contact program in the context of the respective study.

3. A few studies (e.g., Sayler, 1969) compared an intervention group with more than one "untreated" control groups (e.g., both a control group with intra-ethnic small group learning activities and a control group with regular learning activities). The simultaneous inclusion of the comparisons of each control group with the same intervention group would produce dependency problems as just described. In order to conservatively test the effectiveness of contact interventions, we decided to systematically choose the control group that is most similar to the intervention group, for instance, the control group that received a placebo treatment that should not improve interethnic relations (e.g., participation in intra-ethnic small group activities).

4. To counteract dependencies that arise when there is more than one posttest for a given intervention-control comparison, we analyzed different time points of post measurement with separate meta-analytic integrations. Regarding this, we differentiate

between two sets[13]: *direct posttest* (less than one month after the end of the intervention) and *delayed posttest* (one month up to one year after the end of the intervention).

Meta-analytic models. We tested hypothesis 1, the expected positive overall effect of contact interventions, under the assumptions of the random effects model of meta-analytic integration (REM; Hedges, 1983; Hedges & Olkin, 1985; Hedges & Vevea, 1998; Raudenbush, 1994, 2009; Viechtbauer, 2004). The REM suggests a multi-level approach. At the first level, the effect size of each individual study—or intervention-control comparison—estimates the study-specific true effect as given by

$$ES_i = \theta_i + \varepsilon_i, \quad (4)$$

where ES_i is the observed effect size of the *i*th study. The observed effect differs from the study-specific true effect size, denoted by θ_i, due to a sampling error, denoted by ε_i. At the second level, the true effects of all relevant studies (i.e., of all studies that satisfy the inclusion criteria) are assumed to be randomly (normally) distributed with a mean that is the average true effect and a variance that characterizes the variability of the true effects (i.e., the heterogeneity) and is denoted by τ^2. The model equation at that level given by

$$\theta_i = \mu_\theta + \xi_i, \quad (5)$$

where μ_θ is the average true effect of the population of relevant studies, the parameter that represents the mean effect in the REM and that is to be estimated. In addition, ξ_i characterizes the difference between the study-specific true effect of the *i*th study and the average of the true effects of all studies, a term that is represented by the variance of the true effects τ^2 when multiple studies are considered. Since the true effects of the relevant and included studies are assumed to be a random sample drawn from the population of the true effects of all theoretically relevant studies and μ_θ is estimated as a parameter of that population, the REM allows unconditional inferences. Therefore, the meta-analytic results

and conclusions are not restricted to the sample of the included studies but can be generalized to the population of studies that (theoretically) satisfy the inclusion criteria. That is, they are also valid for studies that are conducted and relevant but not included, could have been conducted and would be relevant as well as for those that will be conducted in the future and will satistfy the inclusion criteria. A priori, we have chosen to apply the REM as its theoretical postulate allows the individual true effects to differ. This is in our opinion more appropriate than the assumption of the competing fixed effects model (FEM; Hedges, 1983, 1994; Konstantopoulos & Hedges, 2009; Viechtbauer, 2004). The FEM holds that the included studies have a common true effect and that the observed effect sizes differ solely because of their sampling errors. In addition, the REM allows to generalize the results to the population of (theoretically) relevant studies, while inferences deriving from the use of the FEM are restricted to the sample of the included studies.

Hypotheses 2 to 5 focus on the values of central study characteristics (e.g., ethnic majority vs. minority). With regard to the incorporation of potential moderators in a meta-analytic integration of research studies, similar to the estimation of an overall effect size parameter, there are two competing approaches: the mixed effects model (MEM; Hedges & Olkin, 1985; Raudenbush, 1994, 2009; Viechtbauer, 2004) and the fixed effects model with moderators (FEMwM; Hedges, 1994; Hedges & Olkin, 1985; Konstantopoulos & Hedges, 2009; Viechtbauer, 2004). A priori, we have decided to use the MEM that transfers the multi-level approach of the REM to the fixed values of one (or more) potential moderator(s). That is, within each value (e.g., ethnic majority), the study-specific true effects are estimated by the corresponding observed effects and are randomly (normally) distributed around a value-specific average true effect size μ_θ (e.g., the average true effect for ethnic majorities) with a variance τ^2 that represents the amount of heterogeneity that cannot be accounted for by the moderator(s). Analogue to our previous decision, we think that the MEM is more appropriate

than the FEMwM that transfers the assumptions of the FEM to the values of the potential moderator(s). Additionally, in contrast to the FEMwM, the MEM allows the generalization of the results to the population of (theoretically) relevant studies within the values of the potential moderator(s).

Meta-analytic procedures. We used a restricted maximum likelihood (REML) estimator of τ^2. As demonstrated by Monte Carlo simulations (Viechtbauer, 2005), this estimator is efficient and has few biases. To gain further insights in the heterogeneity of effects, we used Cochran's *Q*-Test for homogeneity (Cochran, 1954; Hedges & Olkin, 1985) as well as the I^2-statistic (Higgins & Thomas, 2002). A significant *Q*-test signals that the observed effect sizes differ more than can be explained by the sampling error as the only variability-generating source. In these cases, it is typically concluded that heterogeneity exists, that is, that the true effects differ. While a significant *Q* indicates heterogeneity, due to low power, a non-significant *Q* does not necessarily imply that the true effects do not differ. The I^2-statistic was introduced by Higgins and Thomas (2002) and expresses the amount of the variability between the observed effect sizes that is due to the variability between the underlying true effects, that is, that is due to heterogeneity.

Hypotheses 2 to 5 were tested by the use of WLS meta-regression models (Steel & Kammeyer-Mueller, 2002; Thompson & Higgins, 2002; Viechtbauer, 2008) under the assumptions of the MEM and with dummy coded predictor variables. We refrained from conducting a common multiple regression test of the hypotheses in each cluster of effect sizes since the hypotheses are of an isolated nature and conceptually independent. In addition, a multiple regression analysis could only be conducted with a reduced sample of intervention-control comparisons. A listwise deletion would limit the number of the included cases since the tests of the different hypotheses are conceptually (see below) based on partially different samples of intervention-control comparisons.

For all procedures, we conducted sensitivity analyses to test the robustness of the results. With these analyses, we investigated the influence of the model choice by analyzing the data also under the rejected FEM and the rejected FEMwM, respectively. Moreover, we examined the impact of potential outliers (i.e., effect sizes that are extreme relative to the others) by using externally standardized residuals (Hedges & Olkin, 1985, Viechtbauer, 2007, Viechtbauer & Cheung, 2010). An effect size with an absolute externally standardized residual that is larger than 1.96 can be regarded as a potential outlier (Viechtbauer, 2007, Viechtbauer & Cheung, 2010). At last, it was tested whether an overestimation of the average true effect resulting from a potential publication bias (Rothstein, Sutton, & Borenstein, 2005) is likely. The best method to rule out such a potential threat to the validity of meta-analytic findings is to conduct comprehensive searches for relevant literature. As already described, we have followed this principle. Consequently, published and unpublished research is included to an approximately equal degree (see below). In light of this beneficial aspect, it is not very likely that the findings of the present meta-analysis are biased. Nevertheless, we utilized procedures to assess whether the results could be affected by a publication bias. First, we tested if the effects of published and unpublished documents differ systematically. In case of no discrepancy, the typically assumed selective publication based on the desirability of the results does not seem to be of a major concern in the content area of the present meta-analysis. Second, funnel plots (Light & Pillemer 1984; Light, Singer, & Willett, 1994) that plot the individual effect sizes against their corresponding standard errors were inspected. An asymmetric distribution of the effect sizes around the estimated average true effect can signal that the sample of the included studies is possibly biased. In this case, there are typically no or few studies in the bottom left corner of the plot (i.e., studies with—compared to the others—low sample and effect sizes and therefore less likely significant results). Third, the

funnel plots were statistically tested for asymmetry with a rank correlation test[14] (Begg & Mazumdar, 1994) and a regression test[15] (Egger, Davey, Smith, Schneider, & Minder, 1997).

All analyses were conducted with the metafor package (Viechtbauer, 2010) for R (R Development Core Team, 2010).

Results

Description of the Included Intervention-Control Comparisons

By means of multiple search strategies we identified 5,568 documents of potential relevance. The abstracts of these documents were read carefully and repeatedly in order to assess if a comprehensive inspection of the full text is indicated. A total of 2,237 documents were discarded because it was clear that at least one inclusion criteria is not satisfied by these papers. The full texts of the remaining 3,331 documents were acquired and read repeatedly with attention to detail. We finally identified 68 documents with 121 independent intervention-control comparisons ($k = 74$ within the primary cluster, $k = 47$ within the secondary cluster) that satisfy all criteria and that were therefore included in our meta-analytic integrations.

Summary statistics of the central characteristics of the included intervention-control comparisons are displayed in Table 1, the references are given in Appendix A, information concerning each intervention-control comparison within the primary and the secondary cluster is given in Appendix B.

Some aspects that are presented in Table 1 are of special interest. First, both in the primary (29.7%) and in the secondary (68.1%) cluster, the last decade (i.e., the time span from 2001 to 2010) is characterized by a greater number of included comparisons than any other decade, indicating that contact interventions are of recent interest. Second, the number of the included published and unpublished comparisons is approximately equal. This holds true for the primary (published: 45.9%, unpublished: 54.1%) as well as for the secondary

(published 53.2%, unpublished: 46.8%) cluster. Third, within the secondary cluster a substantial amount of the included research was conducted in the context of an intractable conflict (59.6%), whereas the corresponding percentage is lower within the primary cluster (12.2%). More specifically, the following (former) conflict settings were included: Cyprus, the Middle East (Jewish and Arab/Palestinian persons), Northern Ireland/Republic of Ireland, and South Africa. Fourth, an inspection of the descriptive results concerning the type of contact intervention reveals that most of the included interventions have implemented direct contact (primary cluster: 82.4%, secondary cluster: 85.1%). In contrast, the number of comparisons that are associated with an indirect contact program is rather low.

Over and above the description delivered in Table 1, some further aspects are of interest. At first, no intervention-control comparison within the primary cluster and just one comparison within the secondary cluster tested the effectiveness of an indirect contact program for ethnic minority members. Moreover, no documents were found that have evaluated an imagined contact intervention in a real-world setting outside the laboratory. In addition, there is no evaluation that contains a delayed posttest that was conducted more than one year after the end of the intervention.

General Effectiveness of Contact Interventions

As stated in hypothesis 1, we expected contact interventions to have a positive impact on ethnic attitudes. We separately assessed the general effectiveness of contact programs within the primary cluster (i.e., with randomized POWC-designs as well as with PPWC-designs) and within the secondary cluster (i.e., with PPSG-designs). Within both clusters, we separately analyzed the impact at the time of the direct posttest (i.e., less than one month after the end of the intervention) and at the time of the delayed posttest (i.e., between one and 12 months after the end of the intervention). The results can be seen in Table 2. Additionally, heterogeneity and sensitivity analyses were conducted.

Results for the primary cluster.

Direct posttest. A total of $k = 74$ intervention-control comparisons including $N = 8{,}656$ participants (intervention: 4,769, control: 3,887) were considered for the meta-analytic integration. The distribution of the individual observed effect sizes is displayed in Figure 1, where they are plotted against their corresponding standard errors. The average true effect is estimated to be $\hat{\mu}_\theta = 0.26$, 95% CI [0.19, 0.33]. The null hypothesis, stating that μ_θ is zero, can be rejected ($z = 7.20$, $p < .001$).

Cochran's *Q*-Test suggests variability among the true effects ($Q = 217.75$, $df = 73$, $p < .001$). The variance of the true effects is estimated to be $\hat{\tau}^2 = 0.05$, 95% CI [0.03, 0.10]. The amount of total variability between the observed effect sizes that is due to heterogeneity is estimated to be $I^2 = 68.47\%$, 95% CI [54.57, 79.64] and can be classified as "moderate to high" (Higgins & Thompson, 2002). The estimated mean true effects of comparisons based on a randomized POWC-design and comparisons based on a PPWC-design do not differ significantly ($Q_{model} = 1.43$, $df = 1$, $p = .23$).

By means of sensitivity analyses, the robustness of the reported results was tested. At first, the data were analyzed under the assumptions of the fixed effects model (FEM), too. The true effect is estimated to be $\hat{\theta} = 0.25$, 95% CI [0.21, 0.28] and differs significantly from zero ($z = 13.79$, $p < .001$). Hence, there are only minor differences between the findings of the two models. Second, the distribution of the individual effect sizes was inspected with regard to potential outliers. Three[16] of the included intervention-control comparisons have an absolute externally standardized residual (2.80, 2.50, 2.00) larger than 1.96 and can therefore be regarded as potential outliers. However, additional case deletion diagnostics showed that meta-analyses without these cases do not result in substantially different estimates of μ_θ and heterogeneity. The respective comparisons were hence not excluded. Finally, it was tested whether it is likely that the results are influenced by a publication bias. Therefore, published

comparisons ($\hat{\mu}_\theta$ = 0.24, 95% CI [0.14, 0.35], k = 34) were directly compared to unpublished comparisons ($\hat{\mu}_\theta$ = 0.27, 95% CI [0.17, 0.37], k = 40). Both categories do not differ significantly (Q_{model} = 0.14, df = 1, p = .71). Moreover, a funnel plot is displayed in Figure 1. As can be seen, the effect sizes do not seem to be distributed asymmetrically around the average true effect. This impression is confirmed by the statistical tests of funnel plot asymmetry. Neither the rank correlation test (Kendall's τ = 0.06, p = .42) nor the regression test (z = 0.29, p = .77) suggests a funnel plot asymmetry.

In summary, the meta-analytic results provide clear evidence for the predicted causal sequence from implemented contact interventions to improved ethnic attitudes at the time of the direct posttest.

Delayed posttest. In order to examine whether the positive impact of contact interventions persists, a delayed posttest was conducted for a sample of k = 6 intervention-control comparisons with N = 985 participants (intervention: 501, control: 484). The average of the true effects is estimated to be $\hat{\mu}_\theta$ = 0.27, 95% CI [0.16, 0.38]. The null hypothesis can be rejected (z = 4.73, p < .001).

Cochran's Q-Test indicates an absence of heterogeneity (Q = 6.93, df = 5, p = .23). The variance of the true effects is estimated to be $\hat{\tau}^2$ = 0.01, 95% CI [0.00, 0.34]. The percentage of the total variability between the observed effect sizes that is due to heterogeneity is estimated to be I^2 = 7.76%, 95% CI [0.00, 93.86] and can be categorized as "low" (Higgins & Thompson, 2002).

Again, sensitivity analyses were conducted. At first, the data were also analyzed under the FEM. An estimated true effect of $\hat{\theta}$ = 0.27, 95% CI [0.17, 0.37] resulted that differs significantly from zero (z = 5.36, p < .001). Moreover, the distribution of the individual effect sizes was inspected with regard to potential outliers. There are no potential outliers since all externally standardized residuals are smaller than 1.96. Lastly, we tested whether the results

are potentially influenced by a publication bias. Published comparisons ($\hat{\mu}_\theta$ = 0.33, 95% CI [0.08, 0.58], k = 4) do not differ (Q_{model} = 0.22, df = 1, p = .63) from unpublished comparisons ($\hat{\mu}_\theta$ = 0.28, 95% CI [-0.03, 0.51], k = 2). In addition, the rank correlation test (Kendall's τ = 0.47, p = .27) as well as the regression test (z = 1.22, p = .22) do not signal an asymmetric funnel plot.

Taken together, our findings demonstrate that the causal link between contact interventions and reduced prejudice still exists when tested with a delay of at least one month after the end of the intervention. Furthermore, the effect at the time of the delayed posttest is of similar size than the effect at time of the direct posttest.

Results for the secondary cluster.

Direct posttest. The meta-analytic integration is based on a total of k = 41 intervention-control comparisons with N = 1,935 participants. The observed effects can be seen in Figure 1. The estimated average true effect is $\hat{\mu}_\theta$ = 0.41, 95% CI [0.32, 0.49] and differs significantly from zero (z = 9.18, p < .001).

Cochran's Q-Test suggests that the true effects differ (Q = 144.35, df = 40, p < .001). The variance of the true effects is estimated to be $\hat{\tau}^2$ = 0.05, 95% CI [0.02, 0.09]. The percentage of the total variability between the observed effect sizes that is due to heterogeneity is estimated to be I^2 = 73.26%, 95% CI [55.76, 83.10] and can be classified as "moderate to high" (Higgins & Thompson, 2002).

By the use of sensitivity analyses, the robustness of the reported results was tested. At first, the data were also analyzed under the assumptions of the refused FEM. The resulting estimate of the true effect is $\hat{\theta}$ = 0.41, 95% CI [0.37, 0.45]. The null hypothesis can be rejected (z = 19.57, p < .001). Both models lead to the same point estimate, the confidence interval, however, is smaller when estimated in line with the assumptions of the FEM. Secondly, the distribution of the individual effect sizes was inspected with regard to potential

outliers. Three[17] of the included intervention-control comparisons have an absolute externally standardized residual (3.22, 2.24, 2.06) larger than 1.96 and therefore they can be regarded as potential outliers. Though, by using additional case deletion diagnostics it became evident that meta-analyses that do not include these cases do not result in substantial different estimates of μ_θ and heterogeneity. Hence, we decided not to exclude the respective comparisons. Finally, it was tested whether the findings could be influenced by a publication bias. Published comparisons ($\hat{\mu}_\theta$ = 0.35, 95% CI [0.23, 0.46], k = 23) were compared to unpublished comparisons ($\hat{\mu}_\theta$ = 0.48, 95% CI [0.35, 0.60], k = 18), the two categories do not differ significantly (Q_{model} = 2.22, df = 1, p = .14). A funnel plot is given in Figure 1. As can be seen, the effect sizes are not distributed asymmetrically around the average true effect. In accordance with that, neither the rank correlation test (Kendall's τ = 0.01, p = .92) nor the regression test (z = -0.46, p = .65) suggests a funnel plot asymmetry.

In summary, we meta-analytically integrated 41 comparisons that are based on the PPSG-design. The results demonstrate that contact interventions have a positive impact on ethnic attitudes when measured less than than one month after the end of the intervention.

Delayed posttest. Delayed posttests were conducted for a sample of k = 17 intervention-control comparisons with a total of N = 464 participants. The average of the true effects is estimated to be $\hat{\mu}_\theta$ = 0.35, 95% CI [0.21, 0.50]. The null hypothesis can be rejected (z = 4.68, p < .001).

Cochran's Q-Test indicates heterogeneity (Q = 46.27, df = 16, p < .001). The variance of the true effects is estimated to be $\hat{\tau}^2$ = 0.05, 95% CI [0.03, 0.62]. The amount of total variability between the observed effect sizes that is due to heterogeneity is estimated to be I^2 = 61.84%, 95% CI [40.00, 89.95] and can be categorized as "medium to high" (Higgins & Thompson, 2002).

Again, by means of sensitivity analyses the robustness of the reported results was tested. First, the data were also analyzed under the FEM. The true effect is estimated to be $\hat{\theta} = 0.34$, 95% CI [0.26, 0.42] and differs significantly from zero ($z = 8.48$, $p < .001$). Second, the distribution of the individual effect sizes was inspected with regard to potential outliers. One of the included intervention-control comparisons[18] has an absolute externally standardized residual (2.58) larger than 1.96 and can therefore be regarded as a potential outlier. However, a meta-analysis without this case does not result in a substantial different estimate of μ_θ and heterogeneity. Therefore, this comparison was not excluded. Third, we tested whether the results could be influenced by a publication bias. Published comparisons ($\hat{\mu}_\theta = 0.23$, 95% CI [-0.06, 0.53], $k = 6$) do not differ significantly ($Q_{model} = 0.83$, $df = 1$, $p = .36$) from unpublished comparisons ($\hat{\mu}_\theta = 0.39$, 95% CI [0.22, 0.56], $k = 11$). The rank correlation test (Kendall's $\tau = 0.00$, $p = 1.0$) as well as the regression test ($z = 0.50$, $p = .61$) do not suggest that the funnel plot is asymmetric.

To summarize, the findings confirm hypothesis 1 both for the primary and the secondary cluster. Interventions that are based on the intergroup contact theory improve ethnic attitudes. In addition, this effect is sustained over time.

A Detailed Look at the Effectiveness of Contact Interventions

Having provided support for hypothesis 1, we now want to present the findings concerning the hypotheses 2 to 5. Due to the small number of comparisons with a delayed posttest, the conducted analyses were limited to the direct posttest. Again, we analyzed the primary and secondary cluster separately. In each cluster, we tested the four hypotheses by the use of WLS meta-regressions. The results are displayed in Table 3.

Ethnic majorities and minorities. We expected status position to be a moderator. Nonetheless, we assumed that contact programs are effective for ethnic minority members, too. In the context of the corresponding analyses, we exclusively considered comparisons that

only include participants who had an ethnic majority status in the implemented contact intervention and comparisons which solely consider persons who had an ethnic minority status in the realized program. Comparisons that include both majority and minority members as well as comparisons that include persons who did not have a status position in reference to the "other" ethnic group(s) in the intervention were not regarded.

As displayed in Table 3, the results of the moderator test in the primary cluster illustrate that contact interventions are more effective (Q_{model} = 4.71, df = 1, p < .05) for ethnic majorities ($\hat{\mu}_\theta$ = 0.37, 95% CI [0.26, 0.48], z = 6.42, p < .001, k = 36) than for ethnic minorities ($\hat{\mu}_\theta$ = 0.16, 95% CI [0.00, 0.31], z = 1.99, p = .047, k = 18). In spite of the significant lower effectiveness, the estimated average true effect for minority participants differs significantly from zero.

In the secondary cluster, comparisons that exclusively include majority members also have a higher estimated average true effect ($\hat{\mu}_\theta$ = 0.46, 95% CI [0.29, 0.63], z = 5.22, p < .001, k = 12) than comparisons that are solely based on minority members ($\hat{\mu}_\theta$ = 0.38, 95% CI [0.22, 0.53], z = 4.65, p < .001, k = 15). However, the difference is not significant (Q_{model} = 0.53, df = 1, p = .47). Again, the impact of contact programs is significant for both majority and minority members.

Taken together, the results largely support hypothesis 2. Contact interventions are more effective for ethnic majorities than for ethnic minorities in the primary cluster. Furthermore, the findings for the primary as well as for the secondary cluster demonstrate that contact programs also have a positive impact on ethnic minority members.

Interventions in the context of intractable conflicts. We hypothesized that contact interventions not only have a positive effect in calm areas, which are free of severe macro-level or societal conflicts between the involved ethnic groups, but also improve ethnic

attitudes in regions with a (former) intractable conflict. For this reason, we analyzed and compared the impact of contact programs in these two settings. Table 3 contains the results.

With regard to the primary cluster, contact interventions that are implemented in the context of an intractable conflict between the involved ethnic groups have an estimated average true effect of $\hat{\mu}_\theta = 0.19$ (95% CI [0.00, 0.38], $z = 1.96$, $p < .05$, $k = 9$), for contact programs which are realized in calm regions the mean of the true effects is estimated to be $\hat{\mu}_\theta = 0.27$ (95% CI [0.19, 0.35], $z = 6.92$, $p < .001$, $k = 65$). As can be seen, both estimated average true effects are positive and significantly different from zero. The two settings do not differ significantly ($Q_{model} = 0.63$, $df = 1$, $p = .43$).

In reference to the secondary cluster, contact programs that are conducted in conflict zones have an estimated mean true effect of $\hat{\mu}_\theta = 0.47$ (95% CI [0.36, 0.59], $z = 7.86$, $p < .001$, $k = 20$), interventions that are realized in the absence of an intractable conflict have an estimated average effect of $\hat{\mu}_\theta = 0.34$ (95% CI [0.21, 0.46], $z = 5.43$, $p < .001$, $k = 21$). While the two estimated mean true effects are significantly different from zero, the difference is not ($Q_{model} = 2.61$, $df = 1$, $p = .11$).

In summary, the findings fully confirm hypothesis 3. Based on the data, we conclude that contact programs not only improve ethnic attitudes in the absence of an intractable conflict, but they also have a positive influence in settings with high-intensity conflicts between different ethnic groups or in those areas that suffered from such a constellation in the recent past.

Type of contact intervention. As stated in hypothesis 4, we expected that both direct and indirect contact interventions improve ethnic attitudes. The results can be seen in Table 3.

Within the primary cluster, the average impact of direct contact interventions is estimated to be $\hat{\mu}_\theta = 0.27$ (95% CI [0.19, 0.34], $z = 6.72$, $p < .001$, $k = 61$), indirect contact

interventions have an estimated mean true effect of $\hat{\mu}_\theta = 0.21$ (95% CI [0.05, 0.38], $z = 2.49$, $p < .05$, $k = 13$). The two effects are respectively significantly different from zero.

In the secondary cluster, direct contact programs have an estimated average true effect of $\hat{\mu}_\theta = 0.43$ (95% CI [0.33, 0.52], $z = 8.84$, $p < .001$, $k = 34$), the mean effect of indirect contact interventions is estimated to be $\hat{\mu}_\theta = 0.33$ (95% CI [0.05, 0.61], $z = 2.88$, $p < .05$, $k = 5$). Again, the impact of both direct and indirect contact programs differs significantly from zero. In addition, within the context of two comparisons, an intervention was implemented that consists of both direct and indirect components. The average effectiveness of this mixture is estimated to be $\hat{\mu}_\theta = 0.24$ (95% CI [-0.12, 0.60], $z = 1.31$, $p = .19$, $k = 2$).

Direct and indirect contact programs do not differ significantly, neither in the primary cluster ($Q_{model} = 0.32$, $df = 1$, $p = .57$) nor in the secondary cluster ($Q_{model} = 1.29$, $df = 2$, $p = .52$).

Taken together, our findings confirm hypothesis 4. The results from both clusters clearly demonstrate that not only direct contact interventions improve ethnic attitudes but also indirect contact programs.

At a more detailed level, we also tested the effectiveness of different sub-types of direct and indirect contact programs. As it is displayed in Table 3, in the primary cluster the average true effect is estimated to be $\hat{\mu}_\theta = 0.28$ (95% CI [0.18, 0.38], $z = 5.49$, $p < .001$, $k = 37$) for contact meetings, $\hat{\mu}_\theta = 0.25$ (95% CI [0.13, 0.38], $z = 4.09$, $p < .001$, $k = 24$) for cooperative learning methods, $\hat{\mu}_\theta = 0.41$ (95% CI [0.17, 0.65], $z = 5.49$, $p < .001$, $k = 7$) for extended contact programs, and $\hat{\mu}_\theta = 0.03$ (95% CI [-0.20, 0.27], $z = 5.49$, $p = .77$, $k = 6$) for virtual contact interventions. Hence, with the exception of virtual contact programs, all other sub-types of contact programs have an estimated mean true effect that differs significantly

from zero. However, it has to be considered that the number of the included comparisons regarding virtual and extended contact is small.

Additional meta-regressions within the two types of contact interventions showed that the sub-types of direct programs do not differ significantly (Q_{model} = 0.10, df = 1, p = .76). In contrast, the two sub-types of indirect contact interventions differ significantly (Q_{model} = 8.46, df = 1, p < .01). Accordingly, extended contact interventions are more effective than virtual contact programs. A comparison of contact meetings, cooperative learning programs, and extended contact did not reveal significant differences (Q_{model} = 1.27, df = 2, p = .53).
It cannot be concluded that extended contact interventions are more effective than contact meetings and cooperative learning methods.

In reference to the secondary cluster, the results concerning the sub-types are identical to the findings previously presented for direct contact, indirect contact, and the combination thereof (see Table 3). All of the included evaluations of direct contact programs tested the impact of contact meetings, all studies on indirect contact realized virtual contact, and the two combined cases implemented contact meetings together with virtual contact. Since, as already presented, indirect contact interventions have a positive estimated average true effect which differs significantly from zero, the secondary cluster provides evidence for the effectiveness of virtual contact programs.

To summarize, based on the reported findings it can be concluded that both direct and indirect contact interventions are effective. When analyzed more closely, a positive impact of contact meetings, cooperative learning methods, and extended contact interventions is verified. The effectiveness of virtual contact programs is not definitely confirmed, a positive effect of that sub-type can only be found in the secondary but not in the primary cluster.

Generalization of intervention effects. Finally, we hypothesized that there is a generalization of the impact of contact interventions to attitudes toward the entire outgroup

(primary transfer effect) and even to attitudes toward outgroups that are not involved in the contact intervention (secondary transfer effect). As can be seen in the final section of Table 1 with reference to the characteristic *DV – level of generalization*, the total sample of intervention-control comparisons can be divided into four categories: comparisons with dependent variables that exclusively measure personal relations with individuals of the target outgroup, that solely capture attitudes toward the entire target outgroup of the intervention, that only measure attitudes toward unspecified ethnic outgroups ("other ethnic groups"), and that cover multiple values of that variable and that are therefore mixed. Although, understandably, none of the included comparisons exclusively focus on outgroups that are not involved in the contact intervention, for two comparisons[19] in the mixed dependent variables category of the primary cluster—among other findings—separate results for instruments which measure attitudes toward such non-target outgroups are reported. To test hypothesis 5, we considered comparisons that fit into the first three categories of the level of generalization. Additionally, we decided to also include the non-target outgroup findings of the two comparisons just mentioned. We did this in order to be able to test a secondary transfer effect in the primary cluster. The other results for these two comparisons as well as the other comparisons with mixed dependent variables were not considered.

Within the primary cluster, intervention-control comparisons that only contain measures at the level of known members of the target outgroup have an average estimated true effect of $\hat{\mu}_0 = 0.26$ (95% CI [0.08, 0.44], $z = 2.83$, $p < .01$, $k = 12$), comparisons that exclusively measured prejudice at the level of the entire target outgroup have an estimated mean impact of $\hat{\mu}_0 = 0.29$ (95% CI [0.18, 0.40], $z = 5.04$, $p < .001$, $k = 33$), the average true effect of comparisons whose variables are located at the more general level of unspecified ethnic outgroups is estimated to be $\hat{\mu}_0 = 0.21$ (95% CI [0.05, 0.37], $z = 2.57$, $p < .01$, $k = 16$), and, finally, the two comparisons with dependent variables at the level of non-target

outgroups have an estimated mean impact of $\hat{\mu}_\theta = 0.26$ (95% CI [-0.20, 0.72], $z = 1.11$, $p = .27$, $k = 2$). The differences between the categories are not significant ($Q_{model} = 0.61$, $df = 3$, $p = .89$). In contrast to the non-target category, the estimated average true effects of the other three categories differ significantly from zero. However, the insignificant finding concerning non-involved outgroups has to be seen in the context of a large standard error as only two comparisons could be considered. The point estimate for this category is comparable to the point estimate of the other categories.

In regard to the secondary cluster, comparisons are included that are exclusively located at the level of the entire target outgroup ($\hat{\mu}_\theta = 0.44$, 95% CI [0.35, 0.54], $z = 8.39$, $p < .001$, $k = 25$) and that solely focus on the measurement of attitudes toward unspecified outgroups ($\hat{\mu}_\theta = 0.37$, 95% CI [0.21, 0.53], $z = 4.45$, $p < .001$, $k = 12$). The two estimated mean true effects differ significantly from zero, whereas the difference between them is not significant ($Q_{model} = 0.59$, $df = 1$, $p = .44$).

Taken together, it can be concluded that the effect of contact interventions is not restricted to an improvement of personal relations toward specific individuals who were involved in the programs. While the demonstrated generalization toward the entire target outgroup confirms the existence of a primary transfer effect in the context of real-world contact interventions, the reported point estimate for attitudes toward non-target outgroups suggests a possible secondary transfer effect of contact programs. Moreover, a further type of generalization is shown by the positive impact of contact interventions at the level of unspecified ethnic outgroups.

Sensitivity analyses. The described tests were also conducted under the assumptions of the fixed effects model with moderators (FEMwM). The findings demonstrated that all hypotheses can be confirmed, too. In addition, a significant moderator effect was found for *intractable conflict vs. no intractable conflict* in the secondary cluster ($Q_{model} = 15.25$, $df = 1$,

$p < .001$), with contact interventions having a higher effect when conducted in (former) conflict zones. Furthermore, the variable *DV – level of generalization* significantly moderates the effect of contact interventions in the secondary cluster ($Q_{model} = 8.46$, $df = 1$, $p < .01$), showing a larger effect for attitudes toward the entire target outgroup than for attitudes toward unspecified ethnic outgroups.

Potential outliers within the tested models were examined by using externally standardized residuals. In summary, there were only small numbers of outliers within the models. The results of the analyses being conducted without these potential outliers do not substantially differ from the reported results. Therefore, we decided not to eliminate the respective comparisons.

Supplementary Results

In addition to our explicitly stated hypotheses, we investigated moderating influences of the further variables that we coded but that were not hypothesized to affect the impact of contact interventions. The respective formal, interventional, and methodological study characteristics are displayed in Table 1. As the other analyses, these tests were conducted separately within the primary and within the secondary cluster. Moreover, within each cluster we conducted separate analyses for different samples of intervention-control comparisons: for the total sample that was used to test hypothesis 3 and 4, for the sub-sample that was used to test hypothesis 2, as well as for the sub-sample that was used to test hypothesis 5. In reference to hypothesis 2, we excluded comparisons that consist of a mixture of majority and minority members and that are based on interventions in which there was no status relation between the involved ethnic groups. With regard to hypothesis 5, we disregarded intervention-control comparisons that involved the measurement of ethnic attitudes at multiple generalization levels.

At first, we analyzed the data of the primary cluster. With the total sample of $k = 74$ comparisons, we found that two variables significantly influence the true effects: mean attrition rate ($\hat{\beta} = -0.59$, $Q_{model} = 4.25$, $df = 1$, $p < .05$), whereas the effectiveness is higher when attrition is lower, and year of publication ($\hat{\beta} = -0.01$, $Q_{model} = 5.52$, $df = 1$, $p < .05$), meaning that the impact is stronger for older comparisons.

Concerning the reduced sample of $k = 54$ comparisons that was used to test hypothesis 2, no further significant moderator besides the majority vs. minority moderator was found.

Regarding the sub-sample of $k = 63$ intervention-control comparisons that we used to test hypothesis 5, year of publication is a moderator ($\hat{\beta} = -0.01$, $Q_{model} = 4.47$, $df = 1$, $p < .05$), again older studies are associated with a higher effectiveness.

Moreover, we examined moderating influences of formal, intervention-based, and methodology-based characteristics in the secondary cluster. In neither the total sample nor the two sub-samples a moderator was found.

Discussion

We meta-analytically evaluated the effectiveness of contact interventions that were implemented in real-life settings in order to improve interethnic relations. The following aspects are worth being discussed.

General Effectiveness of Contact Interventions

The findings clearly demonstrate that contact programs do improve ethnic attitudes. The point estimate of the average true effect is $\hat{\mu}_\theta = 0.26$ in the primary cluster (i.e., when data originating from rigorous evaluation designs are analyzed) and $\hat{\mu}_\theta = 0.41$ in the secondary cluster (i.e., when pre-post changes are integrated that originate from studies without a control group). The difference between the clusters seems plausible since the comparisons in the primary cluster estimate the intervention effect and the comparisons in the

secondary cluster estimate the intervention effect plus a potential time effect. Both point estimates can be classified as “small” to “medium” (Cohen, 1988) and qualify as “educationally relevant” (Tallmadge, 1977). To illustrate the effect in the primary cluster, on a scale with a standard deviation of two, an average participant of a contact intervention is predicted to score about 0.5 scale points (i.e., 26% of the standard deviation) better than an average control group member. In the secondary cluster, an average person is estimated to score about 0.8 scale points (i.e., 41% of the standard deviation) better after having participated in a contact program than before.

Two further aspects are worth mentioning. First, the described positive impact of interethnic contact interventions is stable over time. Our analyses of the delayed posttests that were conducted between one and 12 months after the end of the programs showed that the estimated mean true effects at the time of the delayed posttest and at the time of the direct posttest are of comparable size. However, the number of comparisons with a delayed post measurement is small. Moreover, in none of the included comparisons a measurement of long-term effects (i.e., more than 12 months after the end of the program) was realized. Second, the findings can be generalized to the entire population of studies—as defined by our inclusion criteria—on the effectiveness of structured contact interventions to improve ethnic attitudes. Our conclusions are not restricted to the sample of the included studies since we have chosen to utilize the REM (random effects model) and have comprehensively searched the literature, so that our sample of studies can be seen as an unbiased random sample drawn from the population we have just mentioned. The latter aspect is, at least partially, supported by the fact that we considered an almost equal number of published and unpublished documents.

Taken together, interventions that are based on the intergroup contact theory and that are implemented in real-world settings improve ethnic attitudes. Our meta-analysis clearly

demonstrates that contact programs are an effective instrument for the reduction of ethnic prejudice. Therefore, they are worth implementing. This conclusion is also very important for political decision makers: Introducing and supporting a contact intervention means to make decisions based on evidence.

A Detailed Look at the Effectiveness of Contact Interventions

In addition to the expected general effectiveness of contact programs, we analyzed the impact of interventions that are based on intergroup contact more closely.

Ethnic majorities and minorities. We tested whether interethnic contact programs are more effective for ethnic majorities than for minorities but have a positive outcome for the latter as well. Prior primarily survey-based research demonstrated that majority groups gain more from contact than minorities (Binder et al., 2009; Tropp & Pettigrew, 2005). In accordance with these findings, we found a significantly higher effect for ethnic majorities in our methodological better primary cluster. In spite of the illustrated discrepancy in the primary cluster, the estimated mean effect for ethnic minority members differs significantly from zero both in the primary and in the secondary cluster. Our conclusion therefore is that contact programs are more effective for ethnic majorities. Nonetheless, they also have a positive impact on the ethnic attitudes of minority members, but for them they do not reach the full potential.

According to a possible explanation (Pettigrew & Tropp, 2011; Tropp & Pettigrew, 2005**)**, minority members experience in contact situations—in contrast to majority members—concerns about being confronted with prejudice and discrimination. This, in turn, is assumed to reduce the potential of contact to improve attitudes toward majority groups. We expect such concerns to be more prominent in the context of physical contact programs than in the context of indirect contact interventions (see also below). Therefore, we postulate that the majority-minority difference in the impact of contact programs is smaller for indirect than

for direct contact interventions. Unfortunately, due to a lack[20] of appropriate studies focusing on the outcome of indirect contact programs for minority members, we were not able to test this assumption.

Contact interventions in the face of intractable conflicts. Additionally, we were interested in answering the question whether contact interventions are not merely effective in calm regions but also in regions with intractable conflicts. The results clearly show that programs that are explicitly based on the intergroup contact theory improve ethnic attitudes in either context. Although survey-based research (e.g., Paolini et al., 2004; Tredoux & Finchilescu, 2010) has already signaled that contact can even be effective under the (sometimes) problematic conditions in (former) conflict zones, the findings of the present meta-analysis are impressive. It can be concluded that programs that introduce structured contact between members of ethnic groups whose recent common history is marked by reciprocal hostilities, structural inequalities, and by severe acts of group-based violence typically have beneficial effects on the participants' stance toward the other side. Therefore, the implementation of contact interventions is advisable even when their framework conditions are anything but optimal.

Direct and indirect contact interventions. We expected that not only direct contact interventions have positive outcomes, but that also indirect contact programs have a positive impact on ethnic attitudes. Direct contact interventions evoke face-to-face interactions between members of the involved ethnicities. In contrast, indirect contact interventions implement variants of interethnic contact that are not based on physical interactions, for instance, in the form of extended contact by presenting stories that systematically display friendships between in- and outgroup members (e.g., Cameron et al, 2006; Cameron & Rutland, 2006). The results demonstrate that direct and indirect contact programs are both effective. Moreover, a closer look at the different sub-types of contact interventions within

the primary cluster revealed that indirect programs merely have a positive impact when they are based on extended contact. The number of the included virtual programs, however, is small. In addition, analyses of the secondary, methodologically weaker cluster showed that virtual contact interventions are also able to reduce ethnic prejudice.

While, at the present time, the positive effect of virtual contact interventions is not definitely confirmed, extended contact interventions can be seen as a sound alternative to direct contact programs, in particular, because this type of contact program can also be implemented in areas with no or only a few ethnic outgroup members. Furthermore, extended contact interventions can typically be realized with less effort and costs as well as under more structured conditions than face-to-face contact interventions. For example, a series of designed stories for children can be utilized repeatedly at will and in various places under structured conditions.

However, the findings should be interpreted with a certain caution. Up to now, the number of evaluations of indirect contact interventions is small.

Generalization of intervention effects. Lastly, we hypothesized that the impact of contact interventions is not restricted to an improvement of interpersonal relations toward those outgroup members who took part in the program. The effect of contact interventions was expected to generalize to the entire target outgroup (primary transfer effect) as well as to outgroups not involved in the intervention (secondary transfer effect). The results clearly demonstrate that the impact of contact-based programs on personal relations and on attitudes toward the target outgroup as a whole is similar. Therefore, the findings support the existence of a primary transfer effect of contact interventions. Moreover, the results show that contact programs also have a positive outcome when measured at the level of unspecified ethnic outgroups. Finally, the data in the primary cluster signal that a secondary transfer effect of contact interventions is likely. This finding, however, has to be handled with care since it is

only based on two comparisons and since the positive point estimate for the average mean effect does not differ significantly from zero.

Directions for Future Research on Intergroup Contact

The results of this meta-analysis are of considerable relevance for the intergroup contact theory and for the research based on it. Our findings provide rigorous evidence for its validity since they support the causal pathway from implemented contact in real-word settings to improved interethnic relations. In doing so, our meta-analysis can answer essential questions concerning the effectiveness of contact interventions. Contact programs improve ethnic attitudes of majority and minority members, in settings with and without an intractable conflict, and when they implement direct or indirect contact. Moreover, their effect is characterized by desired generalizations and is stable over time.

In addition, we also revealed some deficits in the current level of knowledge on intergroup contact and contact-based interventions. Therefore, the following impulses for future research can be given.

First, it has to be clarified why contact programs are generally less effective for ethnic minorities than for ethnic majorities. As soon as research is able to give a clear answer to this question, the respective explanation can be directly addressed in contact interventions and future programs can be designed in a way that might improve their positive impact on ethnic minority members.

Second, and associated with the first aspect, the impact of indirect contact programs on ethnic minorities has to be tested. At the present time, it seems plausible that the reduced effect for minority members is connected with concerns evoked by the contact situation (Pettigrew & Tropp, 2011; Tropp & Pettigrew, 2005). If this applies, extended contact interventions might be an alternative that is less obstructive than physical contact, in particular at the initial contact stages. Accordingly, these programs may improve the

effectiveness of contact experiences for minorities. In reference to this, a stepwise approach that starts with extended contact and then also implements physical contact might be particularly beneficial (see Pettigrew, 1998). Positive experiences in the context of initial extended contact may help to avoid that concerns are evoked by the following face-to-face contact.

Third, variants of imagined contact interventions have to be evaluated in applied settings. The results from laboratory studies are promising.

Fourth, in order to be able to draw final conclusions concerning the existence of a secondary transfer effect, future evaluations of contact programs should also include measures of attitudes toward non-target outgroups.

Fifth, evaluations of contact programs should also focus on actual behavior as well as on implicit attitudes. The corresponding findings would help to gain further insights in the effectiveness of contact interventions.

Finally, further meta-analyses in the field of contact programs are needed. Future reviews should examine the effect of contact interventions in the context of other intergroup configurations, for example, with respect to relations between handicapped and non-handicapped persons, between persons of different generations, or between persons of differing sexual orientations.

Implications for Policy and Practice

The clear message of this quantitative review is that contact interventions improve ethnic attitudes. This effect exists for ethnic majorities and minorities, is present for contexts with and without an intractable conflict as well as well as observable for different types of contact interventions. Moreover, contact programs do not only improve attitudes toward individual outgroup members involved in the intervention but also toward the entire target

outgroup, toward unspecified ethnic outgroups and probably also toward non-target outgroups. In addition, the effect is sustained over time.

The evidence that is provided by the present research synthesis—and by other research syntheses—should be perceived, well-understood, and finally be considered by policy makers. Therefore, it is of great importance to transport the results of research that is directly relevant for the solving of social problems into the political arena and to improve its usage. Unfortunately, the findings of previous studies on the usage of scientific results by policy makers (e.g., Weiss, 1987, 1998) are rather disappointing. Hence, the question arises: What can be done to promote that scientific evidence—for example, in the form of short summaries of meta-analytic results—is finally considered by policy makers? We hope that answers to this question will be found, so that decisions can be routinely made on the basis of the available evidence and interventions are implemented that really can help, for instance, contact programs.

Conclusion

The present meta-analysis is the first that systematically evaluates the impact of theory-driven contact interventions that are realized in real-world settings in order to improve relations between different ethnic groups.

The findings clearly demonstrate that structured contact programs are effective. This statement also applies for ethnic minority members, contact in conflict zones, indirect contact interventions, attitudes toward the involved outgroup as a whole, attitudes toward unspecified ethnic outgroups, as well as probably for attitudes toward non-target outgroups.

Nonetheless, future research is needed to further expand the knowledge on the intergroup contact theory as well as on the effectiveness of contact interventions.

References

Acuff, D. S. (1975). *The effect of an outdoor education experience on the general and intercultural anxiety of Anglo and Black sixth-graders* (Unpublished doctoral dissertation). University of Southern California.

Albarracín, D., Gillette, J. C., Earl, A.N., Glasman, L. R., Durantine, M. R., & Ho, M. H. (2005). A test of major assumptions about behavior change: A comprehensive look at the effects of passive and active HIV-prevention interventions since the beginning of the epidemic. *Psychological Bulletin*, *131*, 856–897.

Alderfer, C. P, Alderfer, C. J., Bell, E. L., & Jones, J. (1992). The race relations competence workshop: Theory and results. *Human Relations*, *45*, 1259–1291.

Allgood, I. (1998). *The development, implementation, and evaluation of a Strategic Prejudice Reduction Framework and its effect on dogmatism levels of college students* (Unpublished doctoral dissertation). Florida Atlantic University.

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

Amichai-Hamburger, Y., & McKenna, K. Y. A. (2006). The contact hypothesis reconsidered: Interacting via the Internet. *Journal of Computer-Mediated Communication, 11*, 825–843.

Amir, Y. (1969). Contact hypothesis in ethnic relations. *Psychological Bulletin, 71*, 319–342.

Amir, Y. (1976). The role of intergroup contact in change of prejudice and race relations. In P. A. Katz, & D. A. Taylor (Eds.), *Towards the elimination of racism* (pp. 245–308). New York, NY: Pergamon.

Amir, Y., & Ben-Ari, R. (1985). International tourism, ethnic contact, and attitude change. *Journal of Social Issues, 41*, 105–115.

Amir, Y., & Garti, C. (1977). Situational and personal influences on attitude change. *International Journal of Intercultural Relations, 1*, 58–75.

Anastasopoulos, P. G. (1992). Tourism and attitude change: Greek tourists visiting Turkey. *Annals of Tourism Research, 19*, 629–642.

Asbrock, F., Christ, O., Hewstone, M., Pettigrew, T. F., & Wagner, U. (in prep.). *Comparing the secondary transfer effect of direct and extended intergroup contact: The generalization of positive attitudes and its limitations.*

Ashmore, R. D. (1969). *Intergroup contact as a prejudice-reduction technique: An experimental examination of the shared-coping approach and four alternative explanations* (Unpublished doctoral dissertation). University of California, Los Angeles.

Azar, E. (1990). *The management of protracted social conflict: Theory and cases.* Aldershot: Dartmouth.

Bar-Natan, I., Rosen, Y., & Salomon, G. (2010). *Does friendship between adversaries generalize?* Unpublished manuscript, University of Haifa.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257–278.

Becker, B. J. (2000). *Multivariate meta-analysis*. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling.* (pp. 499–525). San Diego, CA: Academic Press.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101.

Berryman-Fink, C. (2006). Reducing prejudice on campus: The role of intergroup contact in diversity education. *College Student Journal, 40*, 511–516.

Bigler, R. S. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues, 55*, 687–705.

Binder, J., Zagefka, H., Brown, R., Funke, F., Kessler, T., Mummendey, A., … Leyens, J.P. (2009). Does contact reduce prejudice or does prejudice reduce contact? A longitudinal test of the contact hypothesis among majority and minority groups in three European countries. *Journal of Personality and Social Psychology, 96*, 843–856.

Black, J. S., & Mendenhall, M. (1990). Cross-cultural training effectiveness: A review and a theoretical framework for future research. *Academy of Management Review, 15*, 113–136.

Blaney, N. T., Stephan, C., Rosenfield, D., Aronson, E., & Sikes, J. (1977). Interdependence in the classroom: A field study. *Journal of Educational Psychology, 69*, 121–128.

Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in Experimental Social Psychology, 37*, 255–343.

Cameron, L., & Rutland, A. (2006). Extended contact through story reading in school: Reducing children's prejudice towards the disabled. *Journal of Social Issues, 62*, 469 –488.

Cameron, L., Rutland, A., Brown, R. J., & Douch, R. (2006). Changing children's intergroup attitudes towards refugees: Testing different models of extended contact. *Child Development, 77*, 1208–1219.

Carl, M. E., & Jones. L. W. (1972). *Some preliminary observations regarding the Minnesota state human relations requirement and its effect on in-service teachers*. Unpublished manuscript, Moorhead State College.

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology, 84*, 851–862.

Christ, O., Hewstone, M., Tausch, N., Wagner, U., Voci, A., Hughes, J., & Cairns, E. (2010). Direct contact as moderator of extended contact effects. *Personality and Social Psychology Bulletin. 36*, 1662-1674.

Clore, G. L., Bray, R. M., Itkin, S. M., & Murphy, P. (1978). Interracial attitudes and behavior at a summer camp. *Journal of Personality and Social Psychology, 36*, 107–116.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101–129.

Cohen, E. G., & Roper, S. S. (1972). Modification of interracial interaction disability: An application of status characteristic theory. *American Sociological Review, 37*, 643–657.

Cohen, E. G., Lockheed, M. E., & Lohman, M. R. (1976). The center for interracial cooperation: A field experiment. *Sociology of Education, 49*, 47–58.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Cohen. J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Connolly, E. (1992). *The Edu-Train project and the reduction of prejudice* (Unpublished master's thesis). Rand Afrikaans University, Johannesburg.

Cook, S. W. (1971). *The effect of unintended interracial contact upon racial interaction and attitude change*. Unpublished manuscript, University of Colorado.

Cook, S. W. (1984). Cooperative interaction in multiethnic contexts. In N. Miller & M. B. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 155–185). Orlando: Academic Press.

Cook, S. W. (1985). Experimenting on social issues: The case of school desegregation, *American Psychologist, 40*, 452–460.

Crisp, R. J., Stathi, S., Turner, R. N., & Husnu, S. (2009). Imagined intergroup contact: Theory, paradigm, and practice. *Social and Personality Psychology Compass, 2*, 1–8.

Crisp, R. J., & Turner, R. N. (2009). Can imagined interactions produce positive perceptions? Reducing prejudice through simulated social contact. *American Psychologist, 64*, 231–240.

D'Agostino, A. C. (1980). *The effect of an outdoor education project on the cross-cultural attitudes of Black, White and Puerto Rican fifth graders* (Unpublished doctoral dissertation). Rutgers the State University of New Jersey - New Brunswick.

D'Andrea, M., Daniels, J., & Heck, R. (1991). Evaluating the impact of multicultural counseling training. *Journal of Counseling and Development, 70*, 143–150.

Denson, N. (2009). Do curricular and co-curricular diversity activities influence racial bias? A meta-analysis. *Review of Educational Research, 79*, 805–838.

Desai-Patel, S. P. (2005). *An evaluation of the interactive problem solving method* (Unpublished doctoral dissertation). Alliant International University, San Diego.

Dessel, A., & Rogge, M. (2008). Evaluation of intergroup dialogue: A review of the empirical literature. *Conflict Resolution Quarterly, 26*, 199–238.

Deutsch, M., & Collins, M. (1951). *Interracial housing: A psychological evaluation of a social experiment.* Minneapolis, MN: University of Minnesota Press.

DeVries, D., Edwards, K., & Slavin, R. (1978). Biracial learning teams and race relations in the classroom: Four field experiments using Teams-Games-Tournaments. *Journal of Educational Psychology, 70*, 356–362.

Dixon, J., Durrheim, K., Tredoux, C., Tropp, L., Clack, B., & Eaton, L. (2010a). A paradox integration? Interracial contact, prejudice reduction, and perceptions of racial discrimination. *Journal of Social Issues, 66*, 401–416.

Dixon, J., Durrheim, K., Tredoux. C., Tropp, L., Clack, B., Eaton, L., & Quayle, M. (2010b). Challenging the stubborn core of opposition to equality: Racial contact and policy attitudes. *Political Psychology, 31*, 831–855.

Dodson, J. P. (1970). *Participation in a bi-racial encounter group: Its relationship to acceptance of self and others, racial attitudes, and interpersonal orientations* (Unpublished doctoral dissertation). Purdue University.

Doubilet, K. (2007). *Evaluation of Peace Child Israel. Assessment of attitude change in Arab & Jewish participants and spectators (2006–2007)*. Unpublished manuscript.

Dovidio, J. F., Gaertner, S. L., & Kawakami, K. (2003). Intergroup Contact: The past, present, and the future. *Group Processes and Intergroup Relations, 6*, 5–21.

Durantini, M. R., Albarracín, D., Earl, A., & Mitchell, A. L. (2006). Conceptualizing the influence of social agents of behavior change: A meta-analysis of the effectiveness of HIV-prevention interventionists for different groups. *Psychological Bulletin, 132*, 212–248.

Durrheim, K. (2010). Racial contact and change in South Africa. *Journal of Social Issues, 66*, 273–288.

Eaton, W. O., & Clore, G. L. (1975). Interracial imitation at a summer camp. *Journal of Personality and Social Psychology, 29*, 316–327.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

European Union Agency for Fundamental Rights (2010). *Annual report on fundamental rights*. Retrieved from http://fra.europa.eu/fraWebsite/research/publications/

annual_report_on_fundamental_rights/ar2010_part2_en.htm

Foerster, S. W. (1981). *The effects of a U.S. educational experience on the traditional cultural values of Libyan students* (Unpublished doctoral dissertation). The University of Texas at Austin.

Ford, W. S. (1986). Favorable intergroup contact may not reduce prejudice: Inconclusive evidence, 1960–1984. *Sociology and Social Research, 70*, 256–258.

Franzen, C. W. (2009). *A qualitative exploration of the use of a community based cultural immersion experience and ethnographic techniques with MFT MA students as part of multicultural training* (Unpublished doctoral dissertation). University of Connecticut.

Geffner, R. A. (1978). *The effects of cooperative and interdependent interactions on the self-esteem, attitudes, interethnic relations, and intraethnic perceptions of elementary school children: A field experiment* (Unpublished doctoral dissertation). University of California, Santa Cruz.

Genesee, F. (1978). Second language learning and language attitudes. *Working papers on Bilingualism, 16*, 19–42.

Gibson, J. L. (2006). Do strong group identities fuel intolerance. Evidence from the South African case. *Political Psychology, 27*, 665–705.

Gibson, J. L., & Claassen, C. (2010). Racial reconciliation in South Africa: Interracial contact and changes over time. *Journal of Social Issues, 66*, 255–272.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339–355). New York, NY: The Russell Sage Foundation.

Gómez, A., & Huici, C. (2008). Vicarious intergroup contact and the role of authorities in prejudice reduction. *The Spanish Journal of Psychology, 11*, 103–114.

Gudykunst, W. B. (1977). *The effects of an intercultural communication workshop on cross-cultural attitudes and interaction* (Unpublished doctoral dissertation). University of Minnesota.

Gudykunst, W. B. (1979). The effects of an intercultural communication workshop on cross-cultural attitudes and interaction. *Communication Education, 28*, 179–187.

Hannon, P. O. (2004). *An investigation of the impact of short-term quality intergenerational contact on children's attitudes toward older adults* (Unpublished doctoral dissertation). The Pennsylvania State University.

Harwood, J., Paolini, S., Joyce, N., Rubin, M., & Arroyo, A. (2011). Secondary Transfer Effects from Imagined Contact. *British Journal of Social Psychology, 50*, 180–189.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107–128.

Hedges, L. V. (1982). Estimation of effect size from a series of independent studies. *Psychological Bulletin, 92*, 490–499.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93*, 388–395.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York, NY: The Russell Sage Foundation.

Hedges, L. V, & Olkin, I (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V, & Vevea, J. L (1998). Fixed- and Random-Effects Models in Meta-Analysis. *Psychological Methods, 3*, 486–504.

Hegarty, W. H. (1975). Changes in students' attitudes as a result of participating in a simulated game. *Journal of Educational Psychology, 67*, 136–140.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

Horowitz, E. L. (1936). The development of attitude toward the Negro. *Archives of Psychology, 194*, 5–47.

Hull, W. F. (1972). Changes in world-mindedness after a cross-cultural sensitivity group experience. *The Journal of Applied Behavioral Science, 8*, 115–121.

Husnu, S., & Crisp, R. J. (2010). Imagined intergroup contact: A new technique for encouraging greater interethnic contact in Cyprus. *Peace and Conflict, 16*, 97–108.

Ioannou, M. (2009). Comparing direct and extended contact in Cyprus. *Journal of European Psychology Students, 1*, 1–10.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book Company.

Johnson, D. W., Johnson, R. T., & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formulation and a meta-analysis of the research. *Review of Educational Research, 53*, 5–54.

Johnson, D. W., Johnson R. T., & Maruyama, G. (1984). Group interdependence and interpersonal attraction in heterogeneous classrooms: A meta-analysis. In N. Miller & M. B. Brewer (Eds.), *Groups in Contact: The Psychology of Desegregation* (pp. 187–212). Orlando, FL: Academic Press.

Johnson, D. W., Johnson, R. T., Tiffany, M., & Zaidman, B. (1984). Cross-ethnic relationships: The impact of intergroup cooperation and intergroup competition. *The Journal of Educational Research, 78*, 75–79.

Katz, I., & Benjamin, L. (1960). Effects of white authoritarianism in biracial work groups. *Journal of Abnormal and Social Psychology, 61*, 448–456.

Katz, I., Goldston, J., & Benjamin, L. (1958). Behavior and productivity in biracial work groups. *Human Relations, 11*, 123–141.

Kelley, C., & Meyers, J. (1995). *The Cross-Cultural Adaptability Inventory.* Minneapolis, MN: National Computer Systems.

Kelley, C., & Meyers, J. (1999). The Cross–Cultural Adaptability Inventory. In S. Fowler & M. Mumford (Eds.), *Intercultural Sourcebook: Cross-Cultural Training Methods* (Vol. 2, pp, 53–70). Yarmouth, ME: Intercultural Press.

Kelly, E. W., Jr. (1971a, April). *Changes in racial prejudice and dogmatism as a result of interracial counselor training experiences.* Paper presented at American Personnel and Guidance Association Convention, Atlantic City, NJ.

Kelly, E. W., Jr. (1971b). *The effects of a cross-cultural counselor training institute on racial prejudice and dogmatism* (Unpublished doctoral dissertation). University of South Carolina.

Kiu, L. (2006). *The extended contact effect: The influence of typicality on intergroup attitudes* (Unpublished thesis for the degree of honours). Simon Fraser University.

Konstantopoulos, S., & Hedges, L. V. (2009). Analyzing effect sizes: Fixed effects models. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 279–295). New York, NY: The Russell Sage Foundation.

Krahé, B., & Altwasser, C. (2006). Changing negative attitudes towards persons with physical disabilities: An experimental intervention. *Journal of Community and Applied Social Psychology, 16*, 59–69.

Kuhlemeier, H., van den Bergh, H., & Melse, L. (1996). Attitudes and achievements in the first year of German language instruction in Dutch secondary education. *The Modern Language Journal, 80*, 494–508.

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

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 1159–1174.

Liebkind, K., & McAllister, A. (1999). Extended contact through peer modeling to promote tolerance in Finland. *European Journal of Social Psychology, 29*, 765–780.

Liebkind, K., Haaramo, J., & Jasinskaja-Lahti, I. (2000). Effects of contact and personality on intergroup attitudes among different professionals. *Journal of Community and Applied Social Psychology, 10*, 171–181.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research.* Cambridge, MA: Harvard University Press.

Light, R. J., Singer, J. D., & Willett, J. B. (1994). Displaying and communicating findings from a meta-analysis. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 439–453). New York, NY: The Russell Sage Foundation.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis (Applied Social Research Methods Series, Vol. 49)*. Thousand Oaks, CA: Sage.

MacGinty, R. (2010). No war, no peace: Why so many peace processes fail to deliver peace. *International Politics, 47*, 145–162.

MacGinty, R., Muldoon, O. T., & Ferguson, N. (2007). No war, no peace: Northern Ireland after the agreement. *Political Psychology, 28*, 1–11.

Malhotra, D., & Liyanage, S. (2005). Long-term effects of peace workshops in protracted conflicts. *Journal of Education, 37*, 391–408.

Mann, J. H. (1959). The effect of inter-racial contact on sociometric choices and perceptions. *Journal of Social Psychology, 50*, 143–152.

Mann, J. H. (1960). The differential nature of prejudice reduction. *Journal of Social Psychology, 52*, 339–343.

Maoz, I., & Ellis, D. G. (2008). Intergroup communication as a predictor of Jewish-Israeli agreement with integrative solutions to the Israeli-Palestinian conflict: The mediating effects of out-group trust and guilt. *Journal of Communication, 58*, 490–507.

Masi, C. M., Chen, H.-Y., Hawkley, L. C., & Cacioppo, J. T. (in press). A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review.*

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 537–560). New York, NY: The Russell Sage Foundation.

Mazziotta, A., Mummendey, A., & Wright, S. C. (2011). Vicarious intergroup contact effects: Applying social-cognitive theory to intergroup contact research [Special Issue]. *Group Processes and Intergroup Relations, 14*, 255-274,.

McAlister, A., Ama, E., Barroso, C., Peters, R., & Kelder, S. (2000). Promoting tolerance and moral engagement through peer modeling. *Cultural Diversity and Ethnic Minority Psychology, 6*, 363–373.

McGarry, J., & O'Leary, B. (1995). *Explaining Northern Ireland: Broken images.* Oxford, UK: Blackwell.

McGregor, J. (1989). *Strategies to reduce racial prejudice in students: A meta-analysis of research* (Unpublished doctoral dissertation). McGill University, Canada.

McGregor, J. (1993). Effectiveness of role playing and antiracist teaching in reducing student prejudice. *Journal of Educational Research, 86*, 215–226.

Merkwan, J. V., & Smith, T. B. (1999). Tolerance and racial identity among foreign sojourners: Testing the contact hypothesis. *Psychological Reports, 85*, 170–175.

Miller, N., & Davidson-Podgorny, G. (1987). Theoretical models of intergroup relations and the use of cooperative teams as an intervention for desegregated settings. In C. Hendrick (Ed.), *Review of personality and social psychology* (pp. 41–67). Newbury Park, CA: Sage.

Mingleton, H. V. (1993). *Using a cooperative learning model to improve cultural attitudes and increase cultural literacy* (Unpublished master's thesis). Nova University.

Mitte, K. (2005). Meta-analysis of cognitive behavioral treatments for generalized anxiety disorder: A comparison with pharmacotherapy. *Psychological Bulletin, 131*, 785–795.

Morris, S. (2008). Estimating effect sizes from pretest-posttest-control group Designs. *Organizational Research Methods, 11*, 364–386.

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

Moskowitz, J. M., Malvin, J. H., Schaeffer, G. A., & Schaps, E. (1983). Evaluation of a cooperative learning strategy. *American Educational Research Journal, 20*, 687–696.

Moskowitz, J. M., Malvin, J. H., Schaeffer, G. A., & Schaps, E. (1985). Evaluation of jigsaw, a cooperative learning technique. *Contemporary Educational Psychology, 10*, 104–112.

Mussen, P. H. (1950). Some personality and social factors related to changes in children's attitudes toward Negroes. *Journal of Abnormal and Social Psychology, 45*, 423–441.

Muthuswamy, N., Levine, T. R., & Gazel, J. (2006). Interaction-based diversity initiative outcomes: An evaluation of an initiative aimed at bridging the racial divide on a college campus. *Communication Education, 55*, 105–121.

Nagda, B. A., Gurin, P., Sorensen, N., Gurin-Sands, C., & Osuna, S. M. (2009). From separate corners to dialogue and action. *Race and Social Problems, 1*, 45–55.

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

Nagda, B. A., & Zúñiga, X. (2003). Fostering meaningful racial engagement through intergroup dialogues. *Group Processes, & Intergroup Relations, 6*, 111–128.

Okoye-Johnson, O. N. (1999). *A meta-analysis of the effects of multicultural education on the racial attitudes of pre-k through grade 12 students: A comparison of curricular intervention and reinforcement dimensions in suburban and urban educational settings* (Unpublished doctoral dissertation). Old Dominion University.

Paluck, E. L., & Green, D. P. (2009). Prejudice Reduction: What works? A critical look at evidence from the field and the laboratory. *Annual Review of Psychology, 60*, 339–367.

Paolini, S., Hewstone, M., Cairns, E., & Voci, A. (2004). Effects of direct and indirect cross-group friendships on judgments of Catholics and Protestants in Northern Ireland: The mediating role of an anxiety-reduction mechanism. *Personality and Social Psychology Bulletin, 30*, 770–786.

Patterson, P. K. (2006). *Effect of study abroad on intercultural sensitivity* (Unpublished doctoral dissertation). University of Missouri - Columbia.

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

Pettigrew, T. F. (2009). Contact's secondary transfer effect: Do intergroup contact effects spread to non-participating outgroups? *Social Psychology, 40*, 55–65.

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

Pettigrew, T. F., & Tropp, L .R. (2011). *When groups meet: The dynamics of intergroup contact*. London, UK: Routledge/Psychology Press.

Pizam, A., Fleischer, A., & Mansfield, Y. (2002). Tourism and social change: The case of Israeli ecotourists visiting Jordan. *Journal of Travel Research, 41*, 177–184.

Ponterotto, J. G., Rieger, B. P., Barrett, A., Harris, G., Sparks, R., Sanchez, C. M., & Magids, D. (1996). Development and initial validation of the Multicultural Counseling Awareness Scale (MCSA). In G. R. Sodowsky, & J. Impara (Eds.), *Multicultural assessment in counseling and clinical psychology* (pp. 247–282). Lincoln, NE: Buros Institute of Mental Measurement.

R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Raudenbush, S. W. (1994). Analyzing effect sizes: Random effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 302–321). New York, NY: The Russell Sage Foundation.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–316). New York, NY: The Russell Sage Foundation.

Reddick-Gibson, F. L. (1999). *An evaluation of the B.U.I.L.D. Rites of Passage Program on the attitudes and behaviors of adolescent African-American males* (Unpublished doctoral dissertation). University of Sarasota.

Risberg, D. F. (1972). *A pilot study of a curriculum for a field study in human and intergroup education* (Unpublished doctoral dissertation). The University of Wisconsin - Madison.

Rosenthal, R., & Rubin, D. B., (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*, 400–406.

Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134*, 223–246.

Rothbart, M. (1996). Category-exemplar dynamics and stereotype change. *International Journal of Intercultural Relations, 20*, 305–321.

Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues, 41*, 81–104.

Rothbart, M., & John, O. P. (1993). Intergroup relations and stereotype change: A social-cognitive analysis and some longitudinal findings. In P. M. Sniderman, P. E. Tetlock, & E. G. Carmines (Eds.), *Prejudice, politics and the American dilemma* (pp. 32–59). Stanford, CA: Stanford University Press.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.

Rouhana, N. N., & Bar-Tal, D. (1998). Psychological dynamics of intractable ethnonational conflicts. The Israeli–Palestinian case. *American Psychologist, 53*, 761–770.

Rouse, D. E. (2001). *An investigation of the outcomes of short-term diversity training* (Unpublished doctoral dissertation). University of Georgia - Athens.

Rzoska, K., & Ward, C. (1991). The effects of cooperative and competitive learning methods on the mathematics achievement, attitudes toward school, self-concepts and friendship choices of Maori, Pakeha and Samoan children. *New Zealand Journal of Psychology, 20*, 17–24.

Sabar, N., Yogev, A., & Alper, Y. (1987, April). *Effect of direct contact with Jews on attitudes of Israeli Arab youth and its implications on designing a new curriculum*.

Paper presented at a round table session of the American Educational Research Association, Washington, DC.

Salomon, G. (2004). Does peace education make a difference in the context of an intractable conflict? *Peace and Conflict: Journal of Peace Psychology, 10*, 257–274.

Salomon, G. (2006). Does peace education really make a difference? *Peace and Conflict: Journal of Peace Psychology, 12*, 37–48.

Salomon, G. (2009). *Recent research findings by the Center for Research on Peace Education.* Unpublished manuscript, University of Haifa.

Santos Rego, M. A., & Moledo, M. D. M. L. (2005). Promoting interculturality in Spain: Assessing the use of the Jigsaw classroom method. *Intercultural Education, 16*, 293–301.

Sayler, R. I. (1969). *An exploration of race prejudice in college students and interracial contact* (Unpublished doctoral dissertation). University of Washington.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs, 72*, 92–115.

Schmid, K., Hewstone, M., Kuepper, B., Zick, A., & Wagner, U. (in prep.). *Secondary transfer effects on intergroup contact: A cross-national comparison in Europe*.

Schofield, J. W. (1995). Improving intergroup relations among students. In J. A. Banks, & C. A. M. Banks (Eds.), *Handbook on multicultural education* (pp. 628–634). New York, NY: Macmillan Publishing.

Seguin, C. A. K. (2002). *Evaluating educational sustainability: Effects of training and program delivery on adults implementing diversity awareness programs for children at 11 Michigan summer camps* (Unpublished doctoral dissertation). Wayne State University.

Shachar, H., & Sharan, S. (1994). Talking, relating, and achieving: Effects of cooperative learning and whole-class instruction. *Cognition and Instruction, 12*, 313–54.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Shook, N. J., & Fazio, R. H. (2008). Interracial roommate relationships: An experimental field test of the contact hypothesis. *Psychological Science, 19*, 717–723.

Singer, D. (1966). *Interracial attitudes of Negro and White fifth-grade children in segregated and unsegregated schools* (Unpublished doctoral dissertation). Columbia University.

Singer, D. (1967). The influence of intelligence and an interracial classroom on social attitudes. In D. Dentler, B. Mackler, & M. Warshauer (Eds.). *The Urban r's: Race relations as the problem in urban education* (pp. 99–116). New York, NY: Praeger.

Slavin, R. E. (1978). Student teams and comparison among equals: Effects on academic performance and student attitudes. *Journal of Educational Psychology, 70*, 532–538.

Sorensen, N., Nagda, B. A., Gurin, P., & Maxwell, K. (2009). Taking a 'hands on' approach to diversity in higher education: A critical-dialogic model for effective intergroup interaction. *Analyses of Social Issues and Public Policy, 9*, 3–35.

Spangenberg, J., & Nel, E. M. (1983). The effect of equal-status contact on ethnic attitudes. *Journal of Social Psychology, 121*, 173–180.

Spanierman, L. B., Neville, H. A., Liao, H., Hammer, J., & Wang, Y. F. (2008). Participation in formal and informal campus diversity experiences: Effects on students' racial democratic beliefs. *Journal of Diversity in Higher Education, 1*, 108–125.

Spearmon, M. L. (1999). *Effects of learning about racial identity and oppression on social work students* (Unpublished doctoral dissertation). University of Washington.

Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96–111.

St. Jean, G. (2007). *Evaluation of intervention programs designed to address high school racial conflicts* (Unpublished doctoral dissertation). George Mason University.

St. John, N. H. (1975). *School desegregation: Outcomes for children.* New York, NY: Wiley.

Stephan, C. W., Renfro, L., & Stephan, W. G. (2004). The evaluation of multicultural education programs: Techniques and a meta-analysis. In W. G. Stephan, & W. P. Vogt (Eds.), *Education programs for improving intergroup relations: Theory, research, and practice* (pp. 227–242). New York, NY: Teachers College Press.

Stephan, W. G. (1999). *Reducing prejudice and stereotyping in schools (Multicultural Education Series).* New York, NY: Teachers College Press.

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

Stevens, R. J., & Slavin, R. E. (1992). *The cooperative elementary school: Effects on students' achievement, attitudes and social relations*. Unpublished manuscript.

Swart, H., Hewstone, M., Christ, O., & Voci, A. (2010). The impact of crossgroup friendships in South Africa: Affective mediators and multi-group comparisons. *Journal of Social Issues, 66*, 309–333.

Tallmadge, G. K. (1977). *The joint dissemination review panel idea book.*Washington, DC: National Institute of Education, and the US Office of Education.

Tam, J. W. Y. (2001). *Fostering interethnic relations at an urban school: Is knowledge forum combined with face-to-face cooperative learning superior to face-to-face cooperative learning alone?* (Unpublished master's thesis). University of Toronto.

Tam, T., Hewstone, M., Kenworthy, J. B., Cairns, E., Marinetti, C., Geddes, L., & Parkinson, B. (2008). Postconflict reconciliation: Intergroup forgiveness and implicit biases in Northern Ireland. *Journal of Social Issues. 64*, 303–320.

Tausch, N., Hewstone, M., Kenworthy, J., Cairns, E., & Christ, O. (2007). Cross-community contact, perceived status differences, and intergroup attitudes in Northern Ireland: The mediating roles of individual-level vs. group-level threats and the moderating role of social identification. *Political Psychology, 28*, 53–61.

Tausch, N., Tam, T., Hewstone, M., Kenworthy, J., & Cairns, E. (2007). Individual-level and group-level mediators of contact effects in Northern Ireland: The moderating role of social identification. *British Journal of Social Psychology, 46*, 541–556.

Tavakoli, M., Hatami, J., & Thorngate, W. (2010). Changing stereotypes in Iran and Canada using computer mediated communication. *Journal of Intercultural Communication, 23*, 5–5.

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*, 1559–1573.

Tredoux, C., & Finchilescu, G. (2010). Mediators of the contact-prejudice relationship amongst South African students on four university campuses. *Journal of Social Issues, 66*, 289–308.

Tropp, L. R., & Pettigrew, T. F. (2005). Relationships between intergroup contact and prejudice among minority and majority status groups. *Psychological Science, 16*, 651–653.

Turner, R. N., Crisp, R. J., & Lambert, E. (2007). Imagining intergroup contact can improve intergroup attitudes. *Group Processes and Intergroup Relations, 10*, 427–441.

Turner, R. N., Hewstone, M., Voci, A., & Vonofakou, C. (2008). A test of the extended intergroup contact hypothesis: The mediating role of intergroup anxiety, perceived

ingroup and outgroup norms, and inclusion of the outgroup in the self. *Journal of Personality and Social Psychology, 95*, 843–860.

U.S. Department of Justice (2009). *Hate crime statistics of the Federal Bureau of Investigation.* Retrieved from http://www2.fbi.gov/ucr/hc2009/index.html

U.S. Department of State (2009). *The country reports on human rights practices.* Retrieved from http://www.state.gov/g/drl/rls/hrrpt/2009/index.htm

Van Laar, C., Levin. S., & Sidanius, J. (2008). Ingroup and outgroup contact: A longitudinal study of the effects of cross-ethnic friendships, dates, roommate relationships and participation in segregated organizations. In U. Wagner, L. Tropp, G. Finchilescu, & C. Tredoux (Eds.), *Improving intergroup relations:Building on the legacy of Thomas F. Pettigrew* (pp. 127–142). Oxford, UK: Blackwell.

Van Laar, C., Levin, S., Sinclair, S., & Sidanius, J. (2005). The effect of university roommate contact on ethnic attitudes and behavior. *Journal of Experimental Social Psychology, 41*, 329–345.

Viechtbauer, W. (2004). *Choosing between the fixed-, random-, and mixed-effects model in meta-analysis: An analysis of existing and new model selection methods* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30*, 261–293.

Viechtbauer, W. (2007, May). *Outlier detection in meta-analysis via studentized deleted residuals.* Paper presented at the Annual Campbell Colloquium, London, England.

Viechtbauer, W**.** (2008). Analysis of moderator effects in meta-analysis. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 471–487). Thousand Oaks, CA: Sage.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*, 112–125.

Wagner, U., & Hewstone, M. (in press). Intergroup contact. In L. R. Tropp (Ed.), *The Oxford handbook of intergroup conflict.* New York, NY: Oxford University Press.

Walker, I., & Crogan, M. (1998). Academic performance, prejudice, and the Jigsaw classroom: New pieces to the puzzle. *Journal of Community & Applied Social Psychology, 8*, 381–393.

Webster, H. (1958). Changes in attitudes during college. *Journal of Educational Psychology, 49*, 109–17.

Weiss, C. H. (1987). The diffusion of social science research to policymakers: An overview. In G. B. Melton, (Ed.), Reforming the law: Impact of child development research (pp. 63–85). New York, NY: Guilford Press.

Weiss, C. H. (1998). Have we learned anything new about the use of evaluations? *American Journal of Evaluation, 19*, 21–33.

Williams, R. M., Jr. (1947). *The reduction of intergroup tensions.* New York, NY: Social Science Research Council.

Wilner, D. M, Walkley, R. P., & Cook, S. W. (1955). *Human relations in interracial housing: A study of the contact hypothesis.* Minneapolis, MN: University of Minnesota Press.

Wilson, S. J., & Lipsey, M. J. (2007). Effectiveness of school-based intervention programs on aggressive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine 33(Suppl. 2)*, 130–143.

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology, 71*, 136–149.

Wright, A. N., & Tolan, J. (2009). Prejudice reduction through shared adventure: A qualitative outcome assessment of a multicultural education class. *Journal of Experiential Education, 32*, 137–154.

Wright, S. C., Aron, A., & Brody, S. M. (2008). Extended contact and including others in the self: Building on the Allport/Pettigrew legacy. In U. Wagner, L. R. Tropp, G. Finchilescu, & C. Tredoux (Eds.), *Improving intergroup relations: Building on the legacy of Thomas F. Pettigrew* (pp. 143–159). Oxford, UK: Blackwell.

Wright, S. C., Aron, A., McLaughlin-Volpe, T., & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology, 73*, 73–90.

Wright, S. C., & Tropp, L. R. (2005). Language and intergroup contact: Investigating the impact of bilingual instruction on children's intergroup attitudes. *Group Processes and Intergroup Relations, 8*, 309–328.

Appendix A

References of the Documents Considered in the Meta-Analytic Integrations

Al Ramiah, A., Hewstone, M., & Little, T. (2008, June). *Using intergroup contact to improve racial integration: An evaluation of a social intervention*. Paper presented at the general meeting of the European Association of Experimental Social Psychology, Opatija, Croatia.

Bar-Natan, I., Rosen, Y., & Salomon, G. (2010). *Does friendship between adversaries generalize?* Unpublished manuscript, University of Haifa.

Boehm, D., Kurthen, H., & Aniola-Jedrzejek, L. (2008). Do international online collaborative learning projects impact ethnocentrism? *E–Learning and Digital Media, 7*, 133–146.

Boulden, W. T. (2006). Youth leadership, racism, and intergroup dialogue. *Journal of Ethnic & Cultural Diversity in Social Work, 15*, 1–26.

Bratt, C. (2008). The jigsaw classroom under test: No effect on intergroup relations evident. *Journal of Community & Applied Social Psychology, 18*, 403–419.

Cameron, L, Rutland, A., & Brown, R. (2007). Promoting children's positive intergroup attitudes towards stigmatized groups: Extended contact and multiple classification skills training. *International Journal of Behavioral Development, 31*, 454–466.

Cameron, L., Rutland, A., Brown, R. J., & Douch, R. (2006). Changing children's intergroup attitudes towards refugees: Testing different models of extended contact. *Child Development, 77*, 1208–1219.

Clark, M. A. (1998). *The effects of a cross-cultural mentoring program and diversity training on preservice teachers' attitudes and beliefs regarding diverse students* (Unpublished doctoral dissertation). University of Florida.

Connolly, E. (1992). *The Edu-Train project and the reduction of prejudice* (Unpublished master's thesis). Rand Afrikaans University, Johannesburg.

Cook, J. E. (2000). *Effect of cooperative learning (student teams-achievement divisions) on African American and Caucasian students' interracial friendships* (Unpublished doctoral dissertation). Louisiana Tech University.

Cookston, R. R. (1973). *Effects of a short-term, intensive, interracial living experience on interracial social distance and attitudes toward interracial issues* (Unpublished doctoral dissertation). East Texas State University.

Derbaum, M. C. (1981). *The effects of multi-ethnic children's stories on the attitudes of White children in grades K-1 and 2-3* (Unpublished doctoral dissertation). University of Pittsburgh.

Dodson, J. P. (1970). *Participation in a bi-racial encounter group: Its relationship to acceptance of self and others, racial attitudes, and interpersonal orientations* (Unpublished doctoral dissertation). Purdue University.

Furuto, S. B., & Furuto, D. M. (1983). The effects of affective and cognitive treatment on attitude change toward ethnic minority groups. *International Journal of Intercultural Relations, 7*, 149–165.

Ganayem, A., Shonfeld, M., Hoter, E., & Walter, J. B. (2010, September). *A longitudinal field experiment among Israeli Jews and Arabs.* Paper presented at the INTER Network Conference, Malta.

Gant, J. L. (1971). *The development and evaluation of a short-term human relations training program to prepare prospective teachers for multi-racial classrooms* (Unpublished doctoral dissertation). The Florida State University.

Gardner, R. C., Kirby, D. M., Smythe, P. C., Dumas, G., Zelman, M., & Bramwell, J. R. (1974). Bicultural excursion programs: Their effects on students' stereotypes, attitudes and motivation. *The Alberta Journal of Educational Research, 20*, 270–277.

Gonzalez, A. (1979). *Classroom cooperation and ethnic balance* (Unpublished doctoral dissertation). University of California, Santa Cruz.

Grant, A. (2006). *The development of global awareness in elementary students through participation in an online cross-cultural project* (Unpublished doctoral dissertation). Louisiana State University and Agricultural & Mechanical College.

Green, D. P., & Wong, J. S. (2009). Tolerance and the contact hypothesis: A field experiment. In E. Borgida (ed.), *The political psychology of democratic citizenship* (pp. 228–246). New York, NY: Oxford University Press.

Hertz-Lazarowitz, R., Kuppermintz, H., & Lang, J. (1998). Arab-Jewish coexistence: Beit Hagafen coexistence programs. In E. Weiner (Ed.), *The handbook of interethnic coexistence* (pp. 565–584). New York, NY: Continuum.

Hoffman, A. J, Wallach, J., & Sanchez, E. (2009). Reducing ethnocentric ideology via multiethnic community service work: Planting Seeds of Hope. *Making connections: Interdisciplinary Approaches to Cultural Diversity, 11*, 40–50.

Horenczyk, G., & Bekerman, Z. (1997). The effects of intercultural acquaintance and structured intergroup interaction on ingroup, outgroup, and reflected ingroup stereotypes. *International Journal of Intercultural Relations, 21*, 71–83.

Hull, W. F. (1972). Changes in world-mindedness after a cross-cultural sensitivity group experience. *The Journal of Applied Behavioral Science, 8*, 115–121.

Jackson, R., Jr. (1998). *The effects of cooperative learning on the development of cross-racial friendships* (Unpublished doctoral dissertation). Lehigh University.

Jin, L., & Erben, T. (2007). Intercultural learning via instant messenger interaction. *CALICO Journal, 24*, 291–311.

Johnson, D. W., & Johnson, R. T. (1982). Effects of cooperative, competitive, and individualistic learning experiences on cross-ethnic interaction and friendships. *The Journal of Social Psychology,118* , 47–58.

Kowalski, K. L. (1998). The impact of vicarious exposure to diversity on preschoolers' emerging ethnic/racial attitudes. *Early Child Development and Care, 146*, 41–51.

Kropiunigg, U., & Pabst, B. (2007). Selbstwertsteigerung und Vorurteilsreduktion bei Jugendlichen eines multiethnischen Peace Camps. *Psychotherapie Forum, 15*, 63–72.

Küchel, J., & Beelmann, A. (2008, July). *Effects of an extended contact intervention on preventing prejudice in elementary school children*. Poster presented at the International Congress of Psychology, Berlin, Germany.

Lazovsky, R. (2007). Educating Jewish and Arab children for tolerance and coexistence in a situation of ongoing conflict: An encounter program. *Cambridge Journal of Education, 37*, 391–408.

London, L. H. (1995). *Kids' College: Enhancing children's understanding and acceptance of cultural diversity* (Unpublished doctoral dissertation). University of South Carolina.

London, L. H., Tierney, G., Buhin, L., Greco, D. M., & Cooper, C. J. (2002). Kids' College: Enhancing children's appreciation and acceptance of cultural diversity. *Journal of Prevention and Intervention in the Community, 24*, 63–78.

Luiz, D., & Krige, P (1985). The effect of social contact between South African White and Colored adolescent girls: A follow-up study. *The Journal of Social Psychology, 125*, 407–408.

Lyras, A. (2007). *Characteristics and psycho-social impacts of an interethnic educational sport initiative on Greek and Turkish Cypriot youth* (Unpublished doctoral dissertation). University of Connecticut.

Mania, E. W., Riek, B. M., Gaertner, S. L., McDonald, S. A., & Lamoreaux, M. J. (2008). *Quantitative evaluation of the Building Bridges for Peace program*. Unpublished manuscript.

Maoz, I. (2000). An experiment in peace: Reconciliation-aimed workshops of workshops of Jewish-Israeli and Palestinian youth. *Journal of Peace Research, 37*, 721–736.

Markowicz, J. A. (2009). *Intergroup contact experience in dialogues on race groups: Does empathy and an informational identity style help explain prejudice reduction?* (Unpublished doctoral dissertation). The Pennsylvania State University.

Miller, K. M. (1994). *The effects of interethnic contact in cooperative learning groups on student attitudes toward race and ethnicity* (Unpublished doctoral dissertation). University of San Francisco.

Nganga, R. W. (2006). *Impact of cross-cultural interaction on counselor trainees' development of cultural empathy and intercultural sensitivity* (Unpublished doctoral dissertation). University of Wyoming.

O'Connor, J., Ruddle, H., & Keenaghan, C. (1992). *A Model of managed co-operation: An evaluation of Co-Operation North's School and Youth Links Scheme*. Dublin: Co-Operation North.

Ohm, R. M. (1987). *Constructing and reconstructing social distance attitudes* (Unpublished doctoral dissertation). Arizona State University.

Oishi, S. S. (1983). *Effects of team-assisted individualization in mathematics on cross-race and cross-sex interactions of elementary school children* (Unpublished doctoral dissertation). University of Maryland College Park.

Oishi, S., Slavin, R. E., & Madden, M. A. (1983, April). *Effects of student teams and individualized instruction on cross-race and cross-sex friendships.* Paper presented at

the Annual Convention of the American Educational Research Association, Montreal, Quebec, Canada.

O'Neill, E. J. (2008). *Intercultural competence development: Implementing International Virtual Elementary Classroom Activities into public schools in the U.S. and Korea* (Unpublished doctoral dissertation). University of Virginia.

Otis, M. D., & Loeffler, D. N. (2005). Changing youths' attitudes toward difference: A community-based model that works. *Social Work with Groups, 28*, 41–64.

Ryan, M. M. (1975). *The effects of a New Brunswick, Canada studies project upon the attitudes of elementary students toward native Indians and themselves* (Unpublished doctoral dissertation). The University of New Brunswick, Canada.

Sayler, R. I. (1969). *An exploration of race prejudice in college students and interracial contact* (Unpublished doctoral dissertation). University of Washington.

Schleien, S. M. (2008). *Confronting the intractable: An evaluation of the Seeds of Peace experience* (Unpublished doctoral dissertation). University of Waterloo, Canada.

Schuitema, J., & Veugelers, W. (in press). Multicultural contacts in education: a case study of an exchange project between different ethnic groups. *Educational Studies*.

Seaman, J., Beightol, J., Shirilla, P., & Crawford, B. (2009). Contact theory as a framework for experiential activities as diversity education: An exploratory study. *Journal of Experiential Education, 32*, 207–225.

Sharan, S., Kussell, P., Hertz–Lazarowitz, R., Bejarano, Y., Raviv, S., & Sharan, Y. (1984). *Cooperative learning in the classroom: Research in desegregated schools*. Hillsdale, NJ: Lawrence Erlbaum.

Slavin, R. E. (1979). Effects of biracial learning teams on cross-racial friendships. *Journal of Educational Psychology, 71*, 381–387.

Slavin, R. E., & Oickle, E. (1981). Effects of cooperative learning teams on student achievement and race relations: Treatment by race interactions. *Sociology of Education, 54*, 174–180.

Slone, M., Tarrasch, R., & Hallis, D. (2000). Ethnic stereotypic attitudes among Israeli children: Two intervention programs. *Merrill-Palmer Quarterly, 46*, 370–389.

Smith, F. T. (1943). *An experiment in modifying attitudes toward the Negro.* New York, NY: Columbia University.

Sorensen, N. A. (2010). The road to empathy: *Dialogic pathways for engaging diversity and improving intergroup relations* (Unpublished doctoral dissertation). University of Michigan.

Tackaberry, S. N. (1980). *Cooperation versus competition: Effects on Hispanic and Anglo children* (Unpublished doctoral dissertation). The American University.

Tavakoli, M., Hatami, J., & Thorngate, W. (2010). Changing stereotypes in Iran and Canada using computer mediated communication. *Journal of Intercultural Communication, 23*, 5–5.

Trubowitz, J. (1969). *Changing the racial attitudes of children.* New York, NY: Praeger.

Vittrup, B., & Holden, G. W. (in press). Exploring the impact of educational television and parent-child discussions on children's racial attitudes. *Analyses of Social Issues and Public Policy.*

Warring, D., Johnson, D. W., Maruyama, G., & Johnson, R. T. (1985). Impact of different types of cooperative learning on cross-ethnic and cross-sex relationships. *Journal of Educational Psychology, 77*, 53–59.

Wayne, E. K. (2008). Is It Just Talk? Understanding and evaluating intergroup dialogue. *Conflict Resolution Quarterly, 25*, 451–478.

Williams, D. H. (1935). *The effects of an interracial project upon the attitudes of Negro and White girls within the Y.W.C.A* (Unpublished master's thesis). Columbia University.

Williams, E. (1972). *Effects of intergroup discussion on social distance and personal space of black and white students* (Unpublished doctoral dissertation). The University of Texas at Austin.

Wilson, M. D. (2006). *Improving intergroup relations through programmed intergroup contact: A model process and its effects* (Unpublished doctoral dissertation). The Pennsylvania State University.

Yablon, Y. B. (2010). Contact as a means of inter-religious engagement: The role of religious culture in peace-building activities. *International Handbooks of Religion and Education, 4*, 497–512.

Yablon, Y. B., & Katz, Y. J. (2001). Internet-based group relations: A high school peace education project in Israel. *Educational Media International, 38*, 175–82.

Appendix B

Characteristics of the Included Intervention-Control Comparisons

B 1

Characteristics of the Intervention-Control Comparisons of the Primary Cluster

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
1.	Al Ramiah et al.	2008	Unpub.	Majority	No conflict	Meeting	Outgroup	PPWC	>18	414	0.033	—	—
2.	Al Ramiah et al.	2008	Unpub.	Minority	No conflict	Meeting	Outgroup	PPWC	>18	468	0.169	—	—
3.	Al Ramiah et al.	2008	Unpub.	Minority	No conflict	Meeting	Outgroup	PPWC	>18	264	0.037	—	—
4.	Boehm et al.	2010	Published	No hierachy	No conflict	Virtual	Unspecified	PPWC	>18	193	0.026	—	—
5.	Boehm et al.	2010	Published	No hierachy	No conflict	Virtual	Unspecified	PPWC	>18	470	0.127	—	—
6.	Bratt	2008	Published	Majority	No conflict	Cooperative	Mixed	PPWC	10–13	68	0.451	—	—
7.	Bratt	2008	Published	Majority	No conflict	Cooperative	Mixed	PPWC	14–18	152	–0.001	174	0.100
8.	Cameron et al.	2007	Published	Majority	No conflict	Extended	Mixed	POWC	5–9	98	0.461	—	—
9.	Cameron et al.	2006	Published	Majority	No conflict	Extended	Mixed	POWC	5–9	123	0.372	—	—
10.	Clark	1998	Unpub.	Mixture	No conflict	Meeting	Unspecified	PPWC	>18	193	0.103	—	—
11.	Cook	2000	Unpub.	Minority	No conflict	Cooperative	individ.	PPWC	10–13	51	0.326	—	—

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
12.	Cook	2000	Unpub.	Majority	No conflict	Cooperative	individ.	PPWC	10–13	205	0.146	—	—
13.	Cookston	1973	Unpub.	Majority	No conflict	Meeting	Unspecified	PPWC	>18	64	–0.027	—	—
14.	Cookston	1973	Unpub.	Minority	No conflict	Meeting	Unspecified	PPWC	>18	47	0.653	—	—
15.	Derbaum	1982	Unpub.	Majority	No conflict	Extended	Outgroup	POWC	5–9	17	0.270	—	—
16.	Derbaum	1982	Unpub.	Majority	No conflict	Extended	Outgroup	POWC	5–9	26	0.898	—	—
17.	Furuto & Furuto	1983	Published	Majority	No conflict	Meeting	Mixed	POWC	>18	63	0.565	—	—
18.	Gant	1971	Unpub.	Majority	No conflict	Meeting	Outgroup	POWC	>18	27	0.576	—	—
19.	Gant	1971	Unpub.	Minority	No conflict	Meeting	Outgroup	POWC	>18	19	–0.026	—	—
20.	Gonzalez	1979	Unpub.	Majority	No conflict	Cooperative	Outgroup	PPWC	14–18	30	0.394	—	—
21.	Gonzalez	1979	Unpub.	Minority	No conflict	Cooperative	Outgroup	PPWC	14–18	22	1.017	—	—
22.	Gonzalez	1979	Unpub.	Minority	No conflict	Cooperative	Outgroup	PPWC	14–18	30	–0.465	—	—
23.	Gonzalez	1979	Unpub.	Majority	No conflict	Cooperative	Outgroup	PPWC	14–18	86	0.029	—	—
24.	Gonzalez	1979	Unpub.	Majority	No conflict	Cooperative	Outgroup	PPWC	14–18	40	0.084	—	—
25.	Gonzalez	1979	Unpub.	Minority	No conflict	Cooperative	Outgroup	PPWC	14–18	107	0.024	—	—
26.	Grant	2006	Unpub.	No hierachy	No conflict	Virtual	Unspecified	PPWC	10–13	46	–0.043	—	—

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
27.	Green & Wong	2009	Published	Majority	No conflict	Meeting	Unspecified	POWC	14–18	54	0.434	—	—
28.	Hertz-Lazarowitz et al.	1998	Published	Majority	Conflict	Meeting	Outgroup	PPWC	14–18	455	–0.131	—	—
29.	Hertz-Lazarowitz et al.	1998	Published	Minority	Conflict	Meeting	Outgroup	PPWC	14–18	471	0.210	—	—
30.	Horenczyk & Bekerman	1997	Published	No hierachy	No conflict	Meeting	Outgroup	PPWC	14–18	220	0.199	—	—
31.	Hull	1972	Published	Majority	No conflict	Meeting	Unspecified	PPWC	>18	62	0.413	62	0.300
32.	Jackson	1998	Unpub.	Majority	No conflict	Cooperative	Individ.	PPWC	10–13	32	0.951	—	—
33.	Jackson	1998	Unpub.	Minority	No conflict	Cooperative	Individ.	PPWC	10–13	64	0.288	—	—
34.	Johnson & Johnson	1982	Published	Mixture	No conflict	Cooperative	Individ.	POWC	10–13	50	0.331	50	0.573
35.	Kowalski	1998	Published	No hierachy	No conflict	Virtual	Outgroup	PPWC	5–9	30	–0.372	—	—
36.	Küchel & Beelmann	2008	Unpub.	Majority	No conflict	Extended	Mixed	PPWC	5–9	76	0.030	76	0.132

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
37.	Lazovsky	2007	Published	Minority	Conflict	Meeting	Outgroup	PPWC	10–13	58	0.046	—	—
38.	Lazovsky	2007	Published	Majority	Conflict	Meeting	Outgroup	PPWC	14–18	59	–0.125	—	—
39.	Markowicz	2009	Unpub.	Majority	No conflict	Meeting	Unspecified	PPWC	>18	22	0.079	—	—
40.	Miller	1994	Unpub.	Majority	No conflict	Cooperative	Unspecified	PPWC	14–18	21	0.031	—	—
41.	Miller	1994	Unpub.	Minority	No conflict	Cooperative	Unspecified	PPWC	14–18	38	0.160	—	—
42.	Nganga	2006	Unpub.	Majority	No conflict	Meeting	Unspecified	PPWC	>18	37	0.884	—	—
43.	O'Neill	2008	Unpub.	No hierachy	No conflict	Virtual	Unspecified	PPWC	10–13	54	–0.083	—	—
44.	O'Neill	2008	Unpub.	No hierachy	No conflict	Virtual	Unspecified	PPWC	10–13	125	0.258	—	—
45.	Oishi	1983	Unpub.	Mixture	No conflict	Cooperative	Individ.	PPWC	10–13	110	0.210	—	—
46.	Oishi et al.	1983	Unpub.	Mixture	No conflict	Cooperative	Individ.	PPWC	10–13	160	0.207	—	—
47.	O'Connor et al.	1992	Published	No hierachy	Conflict	Meeting	Outgroup	PPWC	14–18	128	0.642	—	—
48.	O'Connor et al.	1992	Published	No hierachy	Conflict	Meeting	Outgroup	PPWC	14–18	126	0.255	—	—
49.	O'Connor et al.	1992	Published	No hierachy	Conflict	Meeting	Outgroup	PPWC	14–18	48	–0.146	—	—
50.	O'Connor et al.	1992	Published	No hierachy	Conflict	Meeting	Outgroup	PPWC	14–18	65	–0.214	—	—
51.	Ryan	1976	Unpub.	Majority	No conflict	Meeting	Outgroup	PPWC	10–13	89	0.957	—	—

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
52.	Ryan	1976	Unpub.	Majority	No conflict	Meeting	Outgroup	PPWC	10–13	127	0.975	—	—
53.	Sayler	1969	Unpub.	Majority	No conflict	Meeting	Outgroup	POWC	>18	35	0.309	—	—
54.	Schuitema & Veugelers	2010	Published	Majority	No conflict	Meeting	Mixed	PPWC	14–18	29	0.363	—	—
55.	Sharan et al.	1984	Published	Minority	No conflict	Cooperative	Outgroup	PPWC	10–13	211	0.428	—	—
56.	Sharan et al.	1984	Published	Majority	No conflict	Cooperative	Outgroup	PPWC	10–13	211	0.685	—	—
57.	Slavin & Oickle	1981	Published	Majority	No conflict	Cooperative	individ.	PPWC	10–13	127	0.280	—	—
58.	Slavin & Oickle	1981	Published	Minority	No conflict	Cooperative	individ.	PPWC	10–13	61	–0.164	—	—
59.	Slavin	1979	Published	Mixture	No conflict	Cooperative	individ.	PPWC	10–13	294	0.340	36	0.751
60.	Slone et al.	2000	Published	Majority	Conflict	Extended	Outgroup	PPWC	10–13	208	0.714	—	—
61.	Smith	1943	Published	Majority	No conflict	Meeting	Outgroup	PPWC	>18	92	0.818	—	—
62.	Sorensen	2010	Unpub.	Mixture	No conflict	Meeting	Unspecified	PPWC	>18	691	0.407	587	0.296
63.	Tackaberry	1980	Unpub.	Mixture	No conflict	Cooperative	individ.	PPWC	10–13	83	0.387	—	—
64.	Trubowitz	1969	Published	Minority	No conflict	Meeting	Mixed	PPWC	10–13	30	0.004	—	—
65.	Trubowitz	1969	Published	Majority	No conflict	Meeting	Mixed	PPWC	10–13	33	0.453	—	—

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
66.	Trubowitz	1969	Published	Majority	No conflict	Meeting	Mixed	PPWC	10–13	33	0.119	—	—
67.	Trubowitz	1969	Published	Minority	No conflict	Meeting	Mixed	PPWC	10–13	33	–0.264	—	—
68.	Vittrup & Holden	2010	Published	Majority	No conflict	Extended	Outgroup	PPWC	5–9	43	0.165	—	—
69.	Warring et al.	1985	Published	Mixture	No conflict	Cooperative	individ.	POWC	10–13	50	0.014	—	—
70.	Williams	1934	Unpub.	Minority	No conflict	Meeting	Outgroup	PPWC	14–18	39	0.010	—	—
71.	Williams	1934	Unpub.	Majority	No conflict	Meeting	Outgroup	PPWC	14–18	36	0.920	—	—
72.	Williams	1973	Unpub.	Minority	No conflict	Meeting	Mixed	PPWC	10–13	96	0.331	—	—
73.	Williams	1973	Unpub.	Majority	No conflict	Meeting	Mixed	PPWC	10–13	96	0.202	—	—
74.	Wilson	2006	Unpub.	Mixture	No conflict	Meeting	Unspecified	PPWC	>18	48	0.139	—	—

Note. Age categories refer to age in years.

B 2

Characteristics of the Intervention-Control Comparisons of the Secondary Cluster

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	*N*	*g*	*N*	*g*
1.	Bar-Natan et al.	2010	Unpub.	Minority	Conflict	Meeting	Outgroup	PPSG	14–18	100	1.072	100	0.648
2.	Bar-Natan et al.	2010	Unpub.	Majority	Conflict	Meeting	Outgroup	PPSG	10–13	110	0.701	110	0.079
3.	Boulden	2007	Published	Mixture	No Conflict	Meeting	Unspecified	PPSG	14–18	202	0.220	—	—
4.	Connolly	1992	Unpub.	Minority	Conflict	Meeting	Outgroup	PPSG	14–18	—	—	11	0.358
5.	Connolly	1992	Unpub.	Minority	Conflict	Meeting	Outgroup	PPSG	10–13	—	—	12	0.567
6.	Connolly	1992	Unpub.	Minority	Conflict	Meeting	Outgroup	PPSG	14–18	—	—	34	0.368
7.	Connolly	1992	Unpub.	Majority	Conflict	Meeting	Outgroup	PPSG	14–18	—	—	61	0.443
8.	Dodson	1970	Unpub.	Majority	No Conflict	Meeting	Outgroup	PPSG	14–18	15	0.550	15	0.553
9.	Dodson	1970	Unpub.	Minority	No Conflict	Meeting	Outgroup	PPSG	14–18	8	0.247	8	–0.001
10.	Ganayem et al.	2010	Unpub.	Majority	Conflict	Virtual	Outgroup	PPSG	>18	18	0.275	—	—
11.	Ganayem et al.	2010	Unpub.	Minority	Conflict	Virtual	Outgroup	PPSG	>18	23	0.224	—	—
12.	Gardner et al.	1974	Published	No hierachy	No Conflict	Meeting	Outgroup	PPSG	14–18	211	0.319	—	—
13.	Hoffman et al.	2009	Published	Mixture	No Conflict	Meeting	Unspecified	PPSG	14–18	10	1.036	—	—

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
14.	Jin & Erben	2007	Published	No hierarchy	No Conflict	Virtual	Unspecified	PPSG	14–18	5	0.526	—	—
15.	Kropiunigg & Pabst	2007	Published	No hierarchy	No Conflict	Meeting	Unspecified	PPSG	14–18	8	0.206	8	0.356
16.	Kropiunigg & Pabst	2007	Published	No hierarchy	No Conflict	Meeting	Unspecified	PPSG	14–18	7	0.330	7	–0.247
17.	Kropiunigg & Pabst	2007	Published	Majority	Conflict	Meeting	Unspecified	PPSG	>18	9	0.094	9	0.083
18.	Kropiunigg & Pabst	2007	Published	Minority	Conflict	Meeting	Unspecified	PPSG	>18	10	0.360	10	0.133
19.	London et al.	2002	Published	Mixture	No Conflict	Meeting	Unspecified	PPSG	14–18	38	0.560	—	—
20.	London	1995	Unpub.	Minority	No Conflict	Meeting	Outgroup	PPSG	14–18	29	0.156	—	—
21.	London	1995	Unpub.	Majority	No Conflict	Meeting	Outgroup	PPSG	14–18	21	0.277	—	—
22.	Luiz & Krige	1985	Published	Majority	Conflict	Meeting	Outgroup	PPSG	14–18	–	–	10	0.509
23.	Luiz & Krige	1985	Published	Minority	Conflict	Meeting	Outgroup	PPSG	10–13	–	–	8	1.204
24.	Lyras	2007	Unpub.	No hierarchy	Conflict	Meeting	Outgroup	PPSG	14–18	32	0.533	—	—

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
25.	Lyras	2007	Unpub.	No hierarchy	Conflict	Meeting	Outgroup	PPSG	14–18	41	0.229	—	—
26.	Mania et al.	2008	Unpub.	Minority	Conflict	Meeting	Mixed	PPSG	>18	13	0.119	—	—
27.	Mania et al.	2008	Unpub.	Majority	Conflict	Meeting	Mixed	PPSG	>18	13	0.698	—	—
28.	Mania et al.	2008	Unpub.	Minority	Conflict	Meeting	Mixed	PPSG	14–18	10	0.419	—	—
29.	Maoz	2000	Published	Majority	Conflict	Meeting	Outgroup	PPSG	14–18	52	0.605	—	—
30.	Maoz	2000	Published	Minority	Conflict	Meeting	Outgroup	PPSG	14–18	48	0.411	—	—
31.	Ohm	1987	Unpub.	Minority	No Conflict	Meeting	Outgroup	PPSG	14–18	86	0.707	—	—
32.	Ohm	1987	Unpub.	Majority	No Conflict	Meeting	Unspecified	PPSG	14–18	149	0.314	—	—
33.	Otis	2005	Published	Mixture	No Conflict	Meeting	Unspecified	PPSG	14–18	96	0.687	—	—
34.	Schleien	2009	Unpub.	Minority	Conflict	Meeting	Outgroup	PPSG	>18	36	0.890	15	1.259
35.	Schleien	2009	Unpub.	Minority	Conflict	Meeting	Outgroup	PPSG	10–13	96	0.351	16	–0.109
36.	Schleien	2009	Unpub.	Majority	Conflict	Meeting	Outgroup	PPSG	14–18	100	0.585	31	0.336
37.	Schuitema & Veugelers	2010	Published	Minority	No Conflict	Meeting	Mixed	PPSG	14–18	21	–0.182	—	—
38.	Seaman et al.	2010	Published	Minority	No Conflict	Meeting	Unspecified	PPSG	14–18	23	0.001	—	—

										Direct posttest		Delayed posttest	
Nr.	Author(s)	Year	Publication	Status	Setting	Intervention	Generalization	Design	Age	N	g	N	g
39.	Seaman et al.	2010	Published	Majority	No Conflict	Meeting	Unspecified	PPSG	14–18	51	0.337	—	—
40.	Tavakoli et al.	2010	Published	No hierarchy	No Conflict	Virtual	Outgroup	PPSG	14–18	15	0.331	—	—
41.	Tavakoli et al.	2010	Published	No hierarchy	No Conflict	Virtual	Outgroup	PPSG	14–18	15	0.414	—	—
42.	Wayne	2008	Published	No hierarchy	No Conflict	Meeting	Outgroup	PPSG	14–18	22	0.160	—	—
43.	Wayne	2008	Published	No hierarchy	No Conflict	Meeting	Outgroup	PPSG	>18	17	0.113	—	—
44.	Yablon	2010	Published	Majority	Conflict	Meeting	Outgroup	PPSG	14–18	41	0.475	—	—
45.	Yablon	2010	Published	Minority	Conflict	Meeting	Outgroup	PPSG	14–18	42	0.672	—	—
46.	Yablon & Katz	2001	Published	Majority	Conflict	Meeting and virtual	Outgroup	PPSG	14–18	46	0.556	—	—
47.	Yablon & Katz	2001	Published	Minority	Conflict	Meeting and virtual	Outgroup	PPSG	14–18	46	–0.062	—	—

Note. Age categories refer to age in years.

[1] Since the number of groups is typically small, a randomization at the group-level cannot be seen as an adequate alternative.

[2] In order to search as broad as possible, we refrained from using target words that are connected with the design of the evaluation studies as many documents do not contain explicit statements concerning the applied evaluation design in their title and abstract.

[3] Within the context of this search component we were assisted by Thomas F. Pettigrew (University of California, Santa Cruz) and Rupert Brown (University of Sussex, UK).

[4] When results of the same study were reported in multiple documents (e.g., in a dissertation thesis as well as in a journal article), we used the document that has the highest publication status (e.g., the journal article) as the primary source for the coding of the formal characteristics.

[5] Due to a lack of sufficient systematic information in the majority of the included documents, an index of implementation quality could not be coded.

[6] In contrast to Hedges's *g*, the calculation of a standard error for Morris's *g* and Becker's *g* technically requires the correlation between the pretest and the posttest. However, for almost all of the included comparisons no information about pre-post correlations is given. Therefore, we decided to set the correlation to .7 which approximates the average test-retest correlations of attitude scales reported in the literature, is also used in other meta-analyses (e.g., Masi, Chen, Hawkley & Cacioppo, 2010), and is a conservative estimate given the few pre-post correlations that are reported in the included studies (median: .88; based on seven documents). Other estimates were examined (.3,.5,.9), the results, however, showed that findings are very similar when another value of the test-retest correlation is utilized.

[7] When a dependent variable was reverse scored (i.e., a lower score indicates a better result), we changed the sign of the calculated effect size.

[9] We assume the absence of an interaction between the factors "intervention vs. control" and "time".

[10] Specialized methods for the integration of dependent effect sizes are offered in the literature (e.g., Gleser & Olkin, 1994; Hedges & Olkin, 1985; Rosenthal & Rubin, 1986). However, the application of these approaches as well as the application of multivariate methods presupposes knowledge of the sample-specific inter-correlations between the different measures. Because these inter-correlations are very rarely reported in the included documents, we have chosen to apply the more conservative technique of calculating ordinary arithmetic means. This procedure was also utilized in the meta-analyses that are cited in the text.

[11] Dependency problems do not arise for studies that include more than one treatment condition with a separate control group for each intervention. In this case, all intervention-control comparisons can be included in the meta-analytic integration as they are based on different samples.

[12] We favour selecting one intervention and not to use an aggregate across intervention conditions because different interventions within one study most often are based on different conceptions.

[13] None of the considered documents did include a long-term test of the effectiveness of contact interventions (i.e., more than one year after the end of the intervention).

[14] The test offered by Begg and Mazumdar (1994) is based on the rank correlation between the standardized effect sizes and the transformed standard errors. A significant rank correlation (Kendall's τ) signals an association between the two variables, in case of funnel plot asymmetry high standard errors should be systematically associated higher effect sizes.

[15] The method provided by Egger et al. (1997) utilizes the inverse of the standard error to predict an index that is calculated as the effect size divided by the corresponding standard error. In case the intercept differs significantly from zero, results could be biased.

[16] Comparison 52 (externally standardized residual: 2.80), comparison 51 (externally standardized residual: 2.50), and comparison 61 (externally standardized residual: 2.00) in table B1.

[17] Comparison 1 (externally standardized residual: 3.22), comparison 37 (externally standardized residual: -2.24), and comparison 47 (externally standardized residual: -2.06) in table B2.

[18] Comparison 34 (externally standardized residual: 2.48) in table B2.

[19] Comparison 17 and comparison 36 in table B1.

[20] In the primary cluster all indirect contact interventions were realized with majority samples and in the secondary cluster only one indirect contact program was evaluated for minority participants.

Table 1

Description of the Included Intervention-Control Comparisons

		Primary cluster (k = 74)		Secondary cluster (k = 47)	
Variable	Value	k	%	k	%
Decade	Before 1961	3	4.1	—	—
	1961 – 1970	5	6.8	2	4.3
	1971 – 1980	17	23.0	1	2.1
	1981 – 1990	11	14.9	4	8.5
	1991 – 2000	16	21.6	8	17.0
	2001 – 2010	22	29.7	32	68.1
Type of document	Published	34	45.9	25	53.2
	Journal article	20	27.0	23	48.9
	Book / book chapter	14	18.9	2	4.3
	Unpublished	40	54.1	22	46.8
	Dissertation / master thesis	35	47.3	15	31.9
	Other unpublished	5	6.8	7	14.9
Country of the first author	USA	50	67.6	18	38.3
	Other	24	32.4	29	61.7
Type of contact intervention	Direct	61	82.4	40	85.1
	Contact meeting	37	50.0	40	85.1
	Cooperative group learning	24	32.4	—	—
	Indirect	13	17.6	5	10.6
	Extended Contact	7	9.5	—	—
	Virtual Contact	6	8.1	5	10.6
	Direct and Indirect	—	—	2	4.3
	Contact meeting and virtual contact	—	—	2	4.3

		Primary cluster (k = 74)		Secondary cluster (k = 47)	
Variable	Value	k	%	k	%
Status	Majority	36	48.6	14	29.8
	Minority	18	24.3	19	40.4
	Majority and minority	9	12.2	4	8.5
	No status hierarchy	11	14.9	10	21.3
Setting	Intractable conflict	9	12.2	26	55.3
	No intractable conflict	65	87.8	21	44.7
Reason for implementation	Research interests	53	71.6	10	21.3
	Practical reasons	20	27.0	37	78.7
	Cannot be specified	1	1.4	—	—
Duration of the intervention (gross time)	1 – 7 Days	6	8.1	16	34.0
	> 1 week – 1 month	10	13.5	18	38.3
	> 1 month – 6 months	48	64.9	7	14.9
	> 2 months – 12 months	4	5.4	6	12.8
	Cannot be specified	6	8.1	—	—
Duration – days with delivery of the intervention	1 – 10 days	30	40.5	32	68.1
	11 – 30 days	20	27.0	7	14.9
	31 – 60 days	6	8.1	2	4.3
	> 60 days	4	5.4	—	—
	Cannot be specified	14	18.9	6	12.8
Duration – net time (i.e., days of delivery multiplied with hours per day)	1 – 10 hours	13	17.6	2	4.3
	11 – 50 hours	35	47.3	15	31.9
	51 – 100 hours	9	12.2	19	40.4
	> 100 hours	4	5.4	3	6.4
	Cannot be specified	13	17.6	8	17.0

		Primary cluster (k = 74)		Secondary cluster (k = 47)	
Variable	Value	k	%	k	%
Country of implementation	Austria	—	—	4	8.5
	Canada	2	2.7	1	2.1
	Cyprus	—	—	2	4.3
	Germany	1	1.4	—	—
	Great Britain	2	2.7	—	—
	Israel	8	10.8	10	21.3
	Malaysia	3	4.1	—	—
	Netherlands	1	1.4	1	2.1
	Northern Ireland/Republic of Ireland	4	5.4	—	—
	Norway	2	2.7	—	—
	USA	45	60.8	21	44.7
	South Africa	—	—	6	12.8
	Mixed	6	8.2	2	4.3
Age of the participants	5 – 9 years	7	9.5	—	—
	10 – 13 years	28	37.8	4	8.5
	14 – 18 years	21	28.4	35	74.5
	> 18 years	18	24.3	8	17.0
Sex of the participants (% female)	0 – 30%	—	—	1	2.1
	31 – 70%	46	62.2	32	68.1
	71 – 100%	7	9.5	10	21.3
	Cannot be specified	21	28.4	4	8.5

		Primary cluster (k = 74)		Secondary cluster (k = 47)	
Variable	Value	k	%	k	%
Design	Posttest only with control (POWC)	11	14.9	—	—
	Pretest-posttest with control (PPWC)	63	85.1	—	—
	Pretest-posttest single group (PPSG)	—	—	47	100.0
Assignment to conditions	Randomized (individuals)	16	21.6	—	—
	Not randomized	58	78.4	—	—
	No control group	—	—	47	100.0
Type of control	No treatment	61	82.4	—	—
	Placebo treatment	13	17.6	—	—
	No control group	—	—	47	100.0
Type of posttests	Only direct (less than 1 month)	68	91.9	30	63.8
	Only delayed (1 – 12 months)	—	—	6	12.8
	Direct and delayed	6	8.1	11	23.4
Interval between the end of the	1 – 7 days	55	74.3	34	82.9
intervention and the direct posttest	> 1 week – less than 1 month	13	17.6	2	4.9
	Cannot be specified	6	8.1	5	12.2
Interval between the end of the	1 month – less than 6 months	4	66.7	12	70.6
intervention and the delayed posttest	6 months – 12 months	2	33.3	5	29.4
Total sample size of the comparison	Up to 30	12	16.2	21	51.2
(direct posttest)	31 – 100	37	50.0	16	39.0
	101 – 250	17	23.0	4	9.8
	251 – 500	7	9.5	—	—
	501 – 750	1	1.4	—	—

		Primary cluster (k = 74)		Secondary cluster (k = 47)	
Variable	Value	k	%	k	%
Mean attrition rate (direct posttest)	Up to 10 %	32	43.2	16	39.0
	11 – 30%	19	25.7	9	22.0
	31 – 50%	10	13.5	2	4.9
	> 50 %			3	7.3
	Cannot be specified	13	17.6	11	26.8
Attrition rate – difference between	Up to 10 %	24	32.4	—	—
the intervention and the control group	11 – 30%	3	4.1	—	—
(direct posttest)	31 – 50%	1	1.4	—	—
	> 50 %	—	—	—	—
	Cannot be specified	46	62.2	—	—
Total number of items measuring	1 – 10	16	21.6	20	42.6
ethnic attitudes	11 – 50	43	58.1	16	34.0
	51 – 100	10	13.5	9	19.1
	> 100	1	1.4	—	—
	Cannot be specified	4	5.4	2	4.3
DV – content	Cognitive	17	23.0	18	38.3
	Affective/Behavioral	22	29.7	6	12.8
	Mixed	35	47.3	23	48.9
DV – type of measure	Likert or social distance scale	39	52.7	36	76.6
	Sociometric	12	16.2	1	2.1
	Semantic differential	2	2.7	1	2.1
	Mixed	21	28.4	9	19.1

		Primary cluster (k = 74)		Secondary cluster (k = 47)	
Variable	Value	k	%	k	%
DV – level of generalization	Known individuals	12	16.2	—	—
	Target outgroup	33	44.6	31	66.0
	Unspecified outgroup	16	21.6	12	25.5
	Mixed	13	17.6	4	8.5

Note. DV = dependent variable.

Table 2

General Effectiveness of Contact Interventions

Cluster	Time of posttest	$\hat{\mu}_\theta$	95% CI	Q	$\hat{\tau}^2$	I^2	k	N
Primary	Direct	0.26***	[0.19, 0.33]	217.75***	0.05	68.47	74	8656
	Delayed	0.27**	[0.16, 0.38]	6.93	0.01	7.76	6	985
Secondary	Direct	0.41***	[0.32, 0.49]	144.35***	0.05	73.26	41	1935
	Delayed	0.35**	[0.21, 0.50]	46.27***	0.05	61.84	17	464

Note. $\hat{\mu}_\theta$ = estimated average of the true effects; CI = confidence interval; Q = homogeneity statistic; $\hat{\tau}^2$ = estimated variance between the true effects; I^2 = amount of true variance among total variance; k = number of intervention-control comparisons; N = total number of participants.

** $p < .01$. *** $p < .001$.

Table 3

Effectiveness of Contact Interventions as a Function of Status, Setting, Type of Contact, Type of Intervention, and Generalization

	Primary cluster				Secondary cluster			
Variable	$\hat{\mu}_\theta$	95% CI	*k*	Q_{model}	$\hat{\mu}_\theta$	95% CI	*k*	Q_{model}
Majority vs. Minority								
Majority	0.37***	[0.26, 0.48]	36	4.71*	0.46***	[0.29, 0.63]	12	0.53
Minority	0.16*	[0.00, 0.31]	18		0.38***	[0.22, 0.53]	15	
Setting								
Intractable conflict	0.19*	[0.00, 0.38]	9	0.63	0.47***	[0.36, 0.59]	20	2.61
No intractable conflict	0.27***	[0.19, 0.35]	59		0.34***	[0.21, 0.46]	21	
Type of contact								
Direct	0.27***	[0.19, 0.34]	61	0.32	0.43***	[0.33, 0.52]	34	1.29
Indirect	0.21*	[0.05, 0.38]	13		0.33*	[0.05, 0.61]	5	
Direct and indirect	—	—	—		0.24	[–0.12, 0.60]	2	
Type of intervention								
Meeting	0.28***	[0.18, 0.38]	37	5.58	0.43***	[0.33, 0.52]	34	1.29
Cooperative	0.25***	[0.13, 0.37]	24		—	—	—	
Extended	0.41***	[0.26, 0.48]	7		—	—	—	
Virtual	0.03	[–0.20, 0.27]	6		0.33*	[0.05, 0.61]	5	
Meeting and virtual	—	—	—		0.24	[–0.12, 0.60]	2	
Generalization								
Individual	0.26**	[0.08, 0.44]	12	0.61	—	—	—	0.59
Target outgroup	0.29***	[0.18, 0.40]	33		0.44***	[0.35, 0.54]	25	
Unspecified outgroup	0.21**	[0.05, 0.37]	16		0.37***	[0.21, 0.53]	12	
Non-target outgroup	0.26	[–0.20, 0.72]	2		—	—	—	

Note. $\hat{\mu}_\theta$ = estimated average of the true effects; CI = confidence interval; *k* = number of intervention-control comparisons; Q_{model} = test whether the average true effects differ between the levels of the moderator.

* $p < .05$. ** $p < .01$. *** $p < .001$.

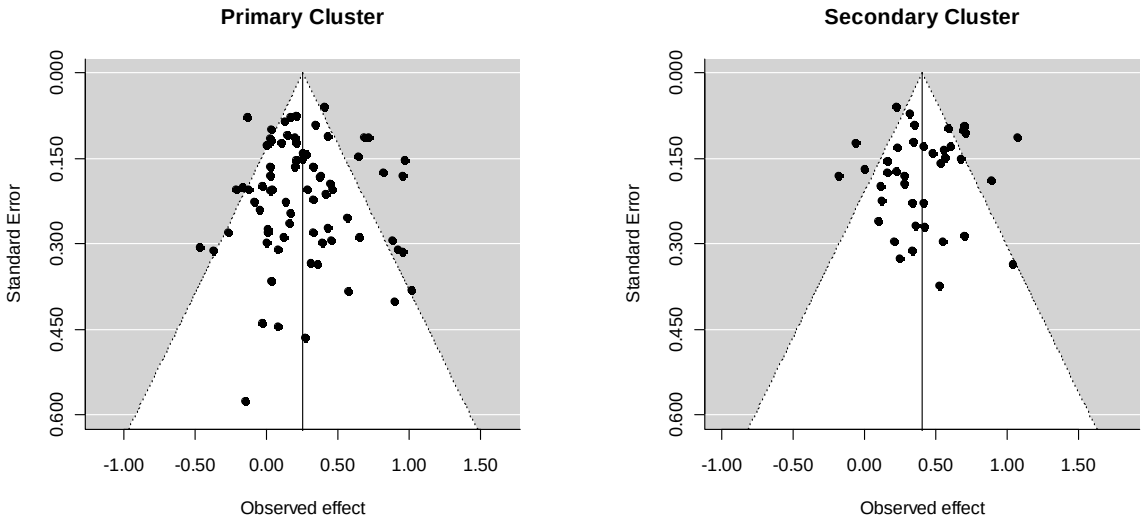

Figure 1. Funnel plots for the direct posttest in the primary ($k = 74$) and secondary ($k = 41$) cluster. The points represent the included intervention-control comparisons. On the axis of abscissae, they display the observed effect sizes, on the axis of ordinates, they display the corresponding standard errors. The estimated average true effect is indicated by a vertical line.

3. Manuscript #2:

The Benefits of Walking in the Shoes of an Outgroup:

A Meta-Analysis of Information Interventions to Improve Ethnic Attitudes

The Benefits of Walking in the Shoes of an Outgroup: A Meta-Analysis of Information Interventions to Reduce Ethnic Prejudice

Gunnar Lemmer and Ulrich Wagner

University of Marburg, Germany

Author Note

Gunnar Lemmer, Department of Psychology, University of Marburg; Ulrich Wagner, Department of Psychology and Center of Conflict Studies, University of Marburg.

This research was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft e.V., WA 1150/9-1/2) awarded to Ulrich Wagner.

Correspondence concerning this article should be addressed to Gunnar Lemmer, Department of Psychology, Section for Psychological Methods, University of Marburg, Gutenbergstr. 18, 35032 Marburg, Germany. Email: gunnar.lemmer@staff.uni-marburg.de.

Abstract

The present meta-analysis tests the effectiveness of information-based interventions that are implemented in order to reduce ethnic prejudice. The term information is used in a broad sense and refers to input that is assumed to improve ethnic attitudes without being based on intergroup contact. In order to exhaustively capture the characteristics of prejudice reduction interventions, we introduce a multi-axial taxonomy encompassing three conceptually independent axes: content, method, and duration. Concerning content, information programs can focus on the enhancement of knowledge, on the evocation of empathy, and/or on the sophistication of social-cognitive skills. In addition, they can utilize passive (e.g., viewing audio-visual material) and/or active (e.g., role plays) methods, can last one or multiple days, and can differ regarding the number of net treatment hours. The meta-analytic test of the general effectiveness of information initiatives is based on a total sample of 154 independent intervention-control comparisons. In line with our prediction, information programs typically improve ethnic attitudes. Interventions that include empathy-evoking content are, as hypothesized, especially effective. Contrary to our expectation, the outcome of initiatives that (also) use active methods does not differ from those that only apply passive techniques. Furthermore, again in opposition to our prediction, the impact of information interventions is not affected by the duration of the treatment. Results, limitations, and directions for future research are discussed.

Keywords: meta-analysis, ethnic attitudes, ethnic prejudice, intervention, knowledge, empathy, social-cognitive skills

The Benefits of Walking in the Shoes of an Outgroup: A Meta-Analysis of Information Interventions to Reduce Ethnic Prejudice

Ethnic prejudice (i.e., negative attitudes toward members of an ethnic outgroup or toward an ethnic outgroup as a whole) is (still) a prevalent social problem (e.g., European Union Agency for Fundamental Rights, 2010; U.S. Department of Justice, 2009; U.S. Department of State, 2009). Psychology provides a substantial number of empirically supported theories (see e.g., Dovidio, Hewstone, Glick & Esses, 2010; Duckitt, 2010; Whitley & Kite, 2006) that can explain the existence of negative ethnic attitudes. However, in order to be able to reduce this social problem, even more important than theoretical explanations is concrete knowledge concerning effective interventions that can diminish prejudicial tendencies.

As meta-analytic research (Lemmer & Wagner, 2011) demonstrates, programs that are explicitly based on the intergroup contact theory (Allport, 1954; see also Pettigrew & Tropp, 2011) have the ability to achieve this goal. In addition, Pettigrew and Tropp (2008; see also Petttigrew & Tropp, 2011) conducted a meta-analysis of primarily survey-based research and found that the effect of contact on intergroup prejudice is mediated by enhanced outgroup-associated knowledge as well as—even stronger—by intensified empathy and reduced anxiety. The paths from knowledge, empathy, and anxiety to prejudice can be interpreted in a general way: Each of these aspects is associated with intergroup attitudes.

Although the available evidence (Lemmer & Wagner, 2011; Petttigrew & Tropp, 2011) unambiguously shows that contact is an effective means to improve intergroup relations, the implementation of contact programs is—for instance in ethnically segregated areas—not always possible. Consequently, the question arises whether there is another way to effectively reduce prejudice. There is reason to be optimistic since the influential mediators of contact found in Pettigrew and Tropp's (2008) meta-analysis cannot only be triggered by

intergroup interactions but also—and probably even more directly—by information-based interventions. In this regard, we use the term *information intervention* in a broad sense to refer to initiatives that are explicitly implemented in order to improve interethnic relations by means of provided input that is assumed to reduce prejudice but not based on intergroup contact.

However, it is not yet clarified whether information interventions—like contact programs—do in fact typically reduce ethnic prejudice. Although several reviews (e.g., Engberg, 2004; McGregor, 1989; 1993; McGregor & Ungerleider, 1993; Paluck & Green, 2009; Stephan & Renfro, 2005; Wagner, van Dick, & Christ, 2002) of strategies to reduce prejudice exist, they do not provide adequate evidence for their effectiveness. Of at least the same importance than the general outcome of information initiatives is the question whether there are particular program components that optimize their impact. Although there is some evidence indicating a beneficial role of empathy (see Batson & Ahmad, 2009), it is poorly understood which features of an intervention enable an optimal reduction of prejudice.

A premise for a fruitful test of the effectiveness of information programs is a taxonomy that is able to systematize them and to extract intervention characteristics. The influence of these characteristics can then be investigated. With regard to this, an adequate typology has to consider the fact that social initiatives are multi-dimensional objects. That is, not only their content (e.g., directed at the enhancement of knowledge or at the evocation of empathy) has to be considered and systematized but also the procedures that are used by them (e.g., rather passive portrayals of information by means of audio-visual material or active involvement of the participants with structured discussions and role plays) as well as their duration, respectively. In addition, these different modalities of interventions have to be represented in a way that conceptually separates and does not mix them up by using just one axis of categorization. Unfortunately, previous suggestions (e.g., Paluck & Green, 2009;

Stephan, 1999) for the systematization of prejudice reduction interventions do not provide such a sophisticated taxonomy.

In view of the major importance of the social problem under consideration, the present paper explicitly deals with essential issues that have not been sufficiently resolved by prior research. First, we present a multi-axial taxonomy of information programs to improve ethnic attitudes. This taxonomy conceptually separates the content of the examined interventions from the methods that are used to pass on the content to the participants and additionally models the duration of the initiatives. Second, the general effect of information programs on ethnic attitudes is tested by means of meta-analytic methods. Third, the influence of content, used methods, and the length of the implemented interventions is meta-analytically examined in order to identify beneficial characteristics and to be able to give concrete recommendations for future activities.

Theories on the Improvement of Ethnic Attitudes by Means of Information

When dealing with interventions that are based on the provision of information, a question immediately arises: Is there theoretical reason to believe that these programs have the ability to effectively reduce interethnic tensions?

Initial indications are given by learning theories. According to the principle of evaluative conditioning (e.g., Martin & Levey, 1978, 1985, 1987; Baeyens & De Houwer, 1995), liking of a stimulus (e.g., an ethnic outgroup) can be induced by systematically pairing the attitude object with positive characteristics. An enormous body of research supports this assumption (e.g., Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). Furthermore, in accordance with the approach of operant conditioning (e.g., Skinner, 1953, 1972), interethnic relations can be improved by systematically rewarding demonstrations of adequate ethnic attitudes and behavior (e.g., Powell-Hopson & Hopson, 1992; Primac 1980). In addition, the social-cognitive learning theory (Bandura, 1963, 1986, 2002) claims that

opinions can be changed and new ways of behaving acquired by means of observing the behavior of a model under several preconditions (e.g., a sufficient degree of attention and an adequate memory capacity of the observer). This statement is confirmed by a substantial amount of research (e.g., Ary, Tildesley, Hops, & Andrews, 1993) and can be directly transferred to the field interethnic relations. Accordingly, observing a model who demonstrates positive ethnic attitudes has the ability to reduce prejudice (e.g., Gopaul-McNicol, 1986; Parakash, 1972).

Learning theories can be regarded as a general approach that illustrates the potential of information-based interventions. In addition, further theories of (ethnic) attitudes (see e.g., Dovidio, Hewstone, Glick & Esses, 2010; Duckitt, 2010; Whitley & Kite, 2006) give reason to expect that information programs may have beneficial effects. Those which have direct relevance to the interventions investigated here refer to three domains: enhancement of knowledge, evocation of empathy, and sophistication of general social-cognitive skills. These approaches will be briefly described hereinafter. As an examination of the available literature revealed that there are no documented information initiatives that have aimed at the improvement of ethnic attitudes by explicitly dealing with the reduction of intergroup anxiety and threat, respectively, we do not address the integrated threat theory of prejudice (Stephan & Stephan, 2000; Stephan, Stephan, & Gudykunst, 1999).

Enhancement of knowledge. Since the middle of the 20th century (e.g., Lasswell, 1948; Hovland, Janis & Kelley, 1953), scholars have been interested in the conceptualization of attitude change by means of communication, that is, if and how attitudes of a receiver can be changed via the provision of "knowledge". With regard to this issue, Petty and Cacioppo (1981, 1986a, 1986b) suggested the elaboration likelihood model (ELM) which is supported by a substantial body of research (e.g., Petty & Wegener, 1998). According to the ELM, providing knowledge—for instance concerning the cultural backgrounds and

accomplishments of ethnic outgroups—can change ethnic attitudes in a strong, stable, and behavior relevant way as long as the receiver is motivated and capable to process the presented arguments. The same prediction is justified by a similar model (heuristic-systematic model; Chaiken, 1980; 1987; Chaiken, Liberman, & Eagly, 1989; Eagly & Chaiken, 1993).

Another line of theorizing gives reason to believe that the promotion of knowledge can reduce prejudice. In accordance with the social identity approach (Tajfel & Turner, 1979, 1986; Turner, Hogg, Oakes, Reicher, & Wetherell, 1989; see also Hogg & Abrams, 1990), a person's memberships in social categories (e.g., European Americans) are relevant to his or her identity. Since humans are motivated to have a positive identity, given salient ethnic group memberships as well as identification with the ethnic ingroup, they distinguish their own group from ethnic outgroups (e.g., African Americans) by evaluating their group more positively. A means to counter this devaluation of other groups is to modify ethnicity-based categorizations via the transmission of new "knowledge". Social psychology offers several empirically supported models (see Gaertner, Dovidio, & Houlette, 2010) to adequately change social categorizations and thereby diminish ethnic prejudice. The de-categorization approach (Brewer & Miller, 1984; Miller, 2002; Wilder, 1978) claims that providing knowledge which characterizes persons as individuals and not as members of groups—and consequently degrading social categorizations—reduces prejudicial tendencies. In contrast, the common ingroup model (Gaertner & Dovidio, 2000; Gaertner, Dovidio, Anastasio, Bachman & Rust; 1993; Gaertner, Mann, Murrell, & Dovidio, 1989; see also Dovidio, Gaertner, Shnabel, Saguy, & Johnson, 2010) advises a group-based re-categorization. Accordingly, emphasizing that the ingroup (e.g., European Americans) and outgroups (e.g., African Americans) have communalities and that both belong to one superordinate group (e.g., Americans) can improve ethnic attitudes.

Evocation of empathy. Emotion-associated processes increasingly gain importance within theories and research on intergroup relations (e.g., Iyer & Leach, 2008; Mackie, Maitner, & Smith, 2010; Mackie, Smith, & Ray, 2008). In this context, *empathy* is of special relevance for ethnic prejudice (e.g., Stephan & Finlay, 1999). Recent conceptualizations of empathy (e.g., Batson & Ahmad, 2009) consider it as an umbrella term that has four inter-related sub-components which can merge: imagining how one would think and feel in an outgroup member's situation (imagine-self perspective), imagining how an outgroup member thinks and feels (imagine-other perspective), feeling as an outgroup member feels (emotion matching), and feeling for an outgroup member (empathetic concern). All of these components are expected to result in more positive intergroup attitudes (Batson & Ahmad, 2009). With special reference to empathetic concern, Batson and colleagues (1997) propose a three-step model of intergroup attitude change by means of the evocation of empathy. First, a person experiences empathetic concern when he or she adopts the perspective of an individual who is a member of a discriminated group and is negatively affected by his or her group membership (e.g., by imagining how this individual's well being is influenced by his or her disadvantaged situation). Second, as a consequence of these empathetic feelings, the person develops a perception of increased concern for the outgroup member's welfare. Finally, as the disadvantaged situation of the individual is associated with his or her group membership, the increased concern generalizes to the entire outgroup and leads to more favorable attitudes toward that group. The general relevance of empathy is demonstrated by a substantial body of research in a highly consistent manner (e.g., Batson et. al, 1997; Batson, Chang, Orr, & Rowland, 2002; Batson et al. 2009) which gives reason to assume that programs that evoke empathy can effectively reduce ethnic prejudice.

Sophistication of social-cognitive skills. A third broad category of theoretical assumptions refers to the sophistication of social-cognitive skills. These skills are of a

general, unspecific nature and not directly related to (ethnic) outgroups. For instance, derived from cognitive development theories (e.g., Aboud, 1988; Bigler & Liben, 1993; Katz, 1976; Piaget & Weil, 1951) it can be expected that the training of classification skills leads to reduced prejudice. Furthermore, some authors (e.g., Eagly, 1987; Schaller, 1994; Schaller & O'Brian, 1992; see also Meiser & Hewstone, 2004), suggest that the formation of erroneous stereotypes can be the consequence of overly simplistic statistical reasoning. However, it is also supposed that this deficit can be reduced by the training of the general cognitive ability to consider potential confounders (e.g., socio-economic status) and thereby to detect spurious correlations (e.g., between ethnicity and performance). An associated approach (e.g., Hamilton & Gifford, 1976; see also Meiser & Hewstone, 2010) stresses the role of the perception of illusory (i.e., non-existing) correlations (Chapman, 1967) for the formation of stereotypes. In this context, it has been demonstrated that persons can incorrectly perceive co-variations when they are repeatedly exposed to two co-occuring but uncorrelated dichotomous variables (e.g., two ethnic groups and two types of behavior) that are respectively skewed distributed. At the same time, there is evidence that such a faulty reasoning can be generally corrected with training (Klauer & Meiser, 2007).

In addition, according to the self-affirmation theory (Steele, 1988; see also Sherman & Cohen, 2006) persons strive to maintain the integrity of the self, that is, a perception of being a good and worthy person. Therefore, threats to the self-integrity motivate to initiate processes that restore positive self-views. One strategy that can be used for that is located at the intergroup level and consists of increased outgroup stereotyping (e.g., Fein & Spencer, 1997) because one's own negative traits are projected to outgroups (Govorun, Fuegen, & Payne, 2006). On the contrary, enhancing self-worth (e.g., via the provision of information) can reduce the likelihood of stereotypic tendencies (e.g., Fein & Spencer, 1997).

To give a general summary, several theoretical approaches give reason to believe that information-based interventions have the ability to reduce ethnic prejudice.

A Multi-Axial Taxonomy of Information Interventions to Reduce Ethnic Prejudice.

Having illuminated possible theoretical backgrounds of information initiatives and, in doing so, provided evidence that these interventions can be expected to be beneficial and are worth further investigation, a second step is to systematize in detail how they can differ. Regarding this issue, a taxonomy that is able to structure the variety of information approaches is a pre-requisite for a detailed test of their effectiveness. Over a time period of more than sixty years, a substantial number of suggestions for the classification of interventions to improve interethnic relations have been made (e.g., Aboud & Levy, 2000; Engberg, 2004, McGregor, 1989; Oskamp, 2000; Pedersen, Walker, Paradies, & Guerin, 2011, Pate 1988, 1995; Pedersen, Walker & Wise, 2005; Rose, 1948; Stephan, 1999; Wagner, Christ, & van Dick, 2002; Williams, 1947). The proposed typologies have great merits and have helped to systematize the field of prejudice reduction programs. However, their utility is limited by two shortcomings. First, these multi-category systems suggest an one-dimensional classification, that is, they differ between several categories (e.g., instructional unit, diversity training, media-based intervention, training of skills) by the use of just one dimension. Therefore, the existing taxonomies do not display whether the interventions that are assigned to different categories address different content with the same or similar procedures (e.g., provision of written material), apply different techniques to pass over the same or similar content (e.g., knowledge about the historic-cultural background of ethnic minorities), or vary regarding content and method. Since prejudice reduction programs are multi-dimensional entities, a one-dimensional schema does not represent them adequately. Second, previous approaches not only project multi-dimensional objects on one dimension but also follow a principle of exclusive (i.e., non-overlapping) classification on that dimension. That is, a given

program is assigned to only one of the defined categories. This practice of "either/or" categorization is highly problematic as a great portion of interventions (e.g., an instructional unit that also use audio-visual material and also train skills) simultaneously covers diverse aspects and cannot be adequately assigned to just one category.

By reason of these circumstances, based on theoretical considerations as well as on an extensive assessment of the available literature on prejudice reduction interventions, we offer a multi-axial taxonomy of information programs (see Figure 1). The proposed classification system encompasses three axes that are conceptually independent of each other: content (i.e., the aspects that are dealt with within the context of an intervention), method (i.e., the ways in which the content is dealt with), and duration (i.e., the length of an intervention).

Axis 1: content. In correspondence with the theoretical approaches already presented, we distinguish between three content components: knowledge-focused, empathy-focused, and focused on social-cognitive skills. Depending on its content, a given intervention can be characterized by just one of these components (e.g., only knowledge-focused), by a combination of two components (e.g., knowledge- and empathy-focused), or by all three components. As will be described, each of the three general content factors, in turn, has three sub-components (see also Figure 1).

Two aspects are worth mentioning before this axis is described in more detail hereinafter. First, we advocate an approach to the classification of information interventions that is focused on the actual content of the programs and not focused on their labels. This is inevitable since terms like "cross-cultural training", "inter-cultural relations course", "diversity training", and "racial awareness workshop" are not used in a consistent fashion. The same label is often used to characterize interventions with different content and programs with very similar content are sometimes labeled with different names. Second,—as already described—we do not consider anxiety in our taxonomy. Even though reduced anxiety is

probably often involved in the process as a by-product of knowledge-focused, empathy-focused, and/or skill-focused content, it is, however, not directly addressed by the content of the present information interventions.

Enhancement of knowledge. With regard to knowledge-based content, we distinguish between three sub-components. A particular program can include none to all of these aspects. The first is termed as *outgroup-focused* and refers to a positive characterization of ethnic outgroups, an explanation of historic-cultural differences, and/or an emphasis of similarities between ethnic outgroups and the ingroup. Second, a further aspect is entitled *prejudice-focused* and is concerned with the transmission of knowledge concerning prejudice and/or discrimination itself. Participants should learn about this phenomenon, how it can be detected, and which general consequences it has. Finally, a third sub-component is *equality-focused* and highlights the social norm of the equality of human beings without focusing on ethnic backgrounds. The three sub-components outgroup-focused, prejudice-focused and equality-focused are closely related to the contents often described with the corresponding traditional terms multi-cultural education (e.g., Banks, 1984, 1995; Grant, 1978; Sleeter, 1985), anti-racism training (e.g., Dei, 1996; Short & Carrington, 1996; Walker, 1989) and colorblind strategies (e.g., Campbell, 1967; Lewis, Chesler, & Forman, 2000; Schofield, 1986). We prefer using the newly introduced terms because the other labels are respectively associated with a certain ideology, while information interventions often realize more than one of the three knowledge-based sub-components (e.g., outgroup- and prejudice-focused).

Evocation of empathy. A second content cluster is not (primarily) concerned with the acquisition of new factual knowledge but rather directly associated with the evocation of empathy by the portrayal of information or experiences initiated by the program. Again, we differentiate between three sub-components and a given intervention can address none to all of them. The first explicitly deals with *personalized cases of ethnic discrimination*. That is,

examples of disadvantaged behavior toward individual members of ethnic outgroups are covered in a personalized fashion. The second sub-component explicitly highlights *feelings of outgroup members* resulting from experienced group-based discriminations (e.g., by means of written material in which minority members describe how they are negatively affected by the way they are typically treated because of their ethnicity). The covered feelings, however, are not negatively directed at the ethnic group of the participants and are not associated with expressions of revenge. Finally, the third sub-component refers to explicit *perspective taking*, that is, the portrayal of material which depicts a member of the ingroup taking the perspective of an ethnic outgroup or exercises which request the participants themselves to take the perspective of a minority group.

It should be noted that whereas the different aspects of empathy that were previously mentioned (i.e., imagine-self perspective, imagine-other perspective, emotion matching, empathetic concern) can be selectively triggered in experiments with specialized instructions (e.g., Batson, et al, 1997), it is typically not possible to systematically separate them in the context of more complex intervention initiatives. Therefore, it cannot be definitely determined which of the four aspects of empathy are triggered by our three content-types and which are not.

Sophistication of social-cognitive skills. Finally, a third content category addresses general social-cognitive skills. These skills are of intra- and inter-individual nature and do not have a special reference to (ethnic) outgroups. Nonetheless, they are expected to influence ethnic attitudes beneficially.

With regard to this component, we also distinguish between three sub-components, a particular program can cover none to all of them. At first, an intervention can deal with the sophistication of *general cognitive skills*, for example by using strategies in order to enhance the complexity of thinking or to promote deductive reasoning. Second, information programs

may deal with the improvement of *general behavioral skills*. For instance, they may train communication skills or conflict management techniques in order to augment these skills. The third sub-component refers to content that explicitly focuses on *self-affirmation* (e.g., by highlighting personal strengths).

To summarize, a given information intervention to reduce ethnic prejudice can include knowledge-based, empathy-focused, and/or skill-associated content. Therefore, seven general orientations of these programs are possible: only knowledge-based; solely empathy-focused; exclusively skills-associated; knowledge-based and empathy-focused; knowledge-based and skill-associated; empathy-focused and skill-associated; knowledge-based, empathy-focused, and skill-associated. If one views each component separately, a particular program can address none to all of the characterized sub-aspects.

Axis 2: method. The content that is covered by an intervention has to be conceptually separated from the methods by which the content is communicated to the participants and/or experienced by them. Therefore, we suggest a second axis of classification that encompasses the utilized procedures. In doing so, we (see also Oskamp, 2000; Oskamp & Schultz, 2005) differentiate between two general types of procedures: passive and active.

Passive methods. Within the context of *passive methods*, the participants are recipients of transmitted information without having an active-creative influence. The following four procedures belong to this group: reading of texts and/or the inspection of visual material, listening to auditive information, viewing of video (i.e., audi-visual) material, and participating in lectures.

Active methods. In contrast to passive procedures, *active methods* give participants the possibility to intervene and to play an active part in the intervention. With regard to this type, we also distinguish between four procedures: structured (small group) discussion, role

play or simulation game, small group activity in order to prepare a "product" (e.g., a poster or a short video), individual activity (e.g., the writing of an essay).

A particular program can apply none to all of the specified passive methods as well as none to all of the listed active elements. Consequently, a given intervention can be assigned to one of three distinct and non-overlapping categories: using only passive methods; applying solely active techniques; employing both passive and active procedures.

Axis 3: duration. Finally, we suggest a third axis that models the duration of information interventions. We propose to consider two different operationalizations of program length. They are conceptually highly correlated but have a different degree of detail. First, we roughly separate one-day programs from multi-day programs, with the latter having a recurring character that one-day programs lack. Second, a more sophisticated—albeit based on available program descriptions not always determinable—conceptualization of treatment length: the number of net intervention hours (calculated as the product of the number of the intervention days and the mean number of the intervention hours per day).

It is important to emphasize that the classification at each of the three axes (i.e., content, method, and duration) is theoretically independent from the categorization at the other two axes. For instance, as previously mentioned, the content perspective taking can be addressed passively by means of presenting audi-visual material showing an ingroup member who takes the perspective of an ethnic minority or actively via role-plays that instruct the participants to take the role of a minority group.

Characteristics not considered. Besides the length of a prejudice reduction initiative, a classification of the intensity of the treatment delivered would be meaningful. However, in contrast to pharmaceutical interventions—where the treatment intensity can be exactly captured in form of milligrams of delivered substance—an adequate and systematic determination of the actual intensity of the provided "substance" is not possible on the basis

of descriptions of social interventions. Furthermore, reviews (e.g., Paluck & Green, 2009) of programs to improve intergroup relations often differentiate between field and laboratory interventions. However, we decided not to incorporate that highly duration-correlated distinction into our taxonomy since these two categories—in contrast to the first impression—frequently cannot be separated accurately. Regarding this issue, it can be additionally assumed that there is no functional difference between programs that, for example, present a particular video in a specialized room at a university or in a room at the school the recipients are attending. Moreover, in contrast to anti-violence programs, initiatives to generate more positive ethnic attitudes can be typically classified as primary and universal prevention strategies (i.e., intervening irrespective of the degree of prejudice of the individual participants). Hence, a differentiation between primary, secondary, and tertiary approaches (Kaplan, 1964) or between universal, selective, and indicated prevention (Gordon, 1983) is also not considered.

Reviews of the Effectiveness of Information Interventions to Improve Ethnic Attitudes

Having provided theoretical evidence for the assumption that information programs are effective and specified how they can be characterized by means of a multi-axial taxonomy, two central issues arise. First, although the presented theoretical approaches give reason to believe that information-based interventions have the ability to change ethnic attitudes, it has to be verified by outcome evaluations whether they are, in fact, effective. Second, in order to optimally counteract the social problem of ethnic prejudice, it has to be scrutinized in reference to the suggested multi-axial taxonomy whether there are certain intervention characteristics that systematically amplify the effectiveness of information initiatives. As already mentioned, several reviews of programs to reduce prejudice exist (e.g., Bigler, 1999; McGregor, 1989, 1993; Engberg, 2004; Paluck & Green, 2009; Pate, 1981; Pedersen, Walker, Paradies, & Guerin, 2011; Pedersen, Walker & Wise, 2005; Stephan,

Renfro, & Stephan, 2004; Stephan & Stephan, 1984; Wagner, Christ, & van Dick, 2002). Can these syntheses provide satisfactory evidence and clear answers?

Content of information interventions. Although the previous reviews typically used an one-dimensional concept of classification and, in doing, so usually merged content and method, their conclusions primarily refer to our content axis. With reference to the impact of interventions that concentrate on the *enhancement of knowledge*, the existing research syntheses deliver contradictory and insufficient findings. The great majority of these reviews applied qualitative procedures or the method of vote-counting (i.e., counting the number of studies with significant positive, negative, and insignificant findings without taking into account the respective sample and effect sizes). Some of them (e.g., Bigler, 1999; Mendenhall, Stahl, Ehnert, Oddou, Osland, & Kühlmann, 2004; Pate, 1981; Stephan & Stephan, 1984) came to the conclusion that the general impact of knowledge-based programs has to be seriously questioned, in particular with regard to their long term effect. Other reviews of the same type (e.g., Banks, 1995; Engberg, 2004; Paluck & Green, 2009; Pedersen, Walker, & Wise, 2005) provide more optimistic findings but also admit that unresolved issues about the effectiveness of information interventions remain.

Meta-analyses that integrate the effect sizes of previous evaluations are typically superior to other types of syntheses. They are guided by more systematic and transparent procedures and deliver more concrete information regarding the general impact of interventions as well as concerning moderating variables. The few existing meta-analyses on prejudice reduction programs (Denson, 2009; McGregor, 1989, 1993; McGregor & Ungerleider, 1993; Okoye-Johnson, 1999; Stephan, Renfro & Stephan, 2004) report positive mean effect sizes, but all of them have major drawbacks. First, they are only based on limited literature searches and small numbers of included evaluations (ranging from 19 to 35 studies). Second, as indicated by their inclusion criteria, they only consider specific types of

interventions and participants. Additionally, they are either not contemporary (McGregor, 1989, 1993; McGregor & Ungerleider, 1993) or do solely incorporate published studies (Stephan, Renfro & Stephan, 2004) and therefore could have upwardly biased results (see Rothstein, Sutton, & Borenstein, 2005). For that reason, the informative value of these analyses is considerably limited.

To sum up, although there is empirically supported reason to believe that the enhancement of knowledge is an effective strategy to reduce prejudice, a definitive answer still needs to be provided.

With regard to *empathy-based components*, Pedersen et al. (2005, 2011) recommend in their reviews that focusing on empathy is very beneficial. In accordance with that, Paluck and Green (2009) conclude that addressing empathy is effective. However, no meta-analysis of empathy-based interventions exists. Due to the absence of meta-analytic findings, a look at the primary research on empathy interventions is indicated. A multiplicity of experiments have been conducted to investigate the impact of empathy on attitudes toward diverse outgroups (e.g., Batson et al. 1997a; Batson et al. 1997b, Batson, Chang, Orr, & Rowlan, 2002; Esses & Dovidio, 2002; Galinsky & Moskowitz, 2000; Finlay & Stephan, 2000; Shih, Wang, Trahan Bucher, & Stotzer, 2009; Vescio, Sechrist, & Paolucci, 2003). These studies demonstrate the suggested effects of empathy in a highly consistent manner.

Taken together, there is substantial evidence suggesting that ethnic prejudice can be reduced by evoking empathy. However, meta-analytic tests of the effect of empathy-associated programs on ethnic attitudes have yet to be provided.

Qualitative syntheses (e.g., Aboud & Levy, 2000) that also review the impact of programs that train *social-cognitive skills* draw positive conclusions. Though, the effect of interventions with skill-promoting content on ethnic attitudes has not yet been scrutinized by using meta-analytic techniques. The findings of primary research demonstrate that the

training of deductive reasoning (e.g., Schaller, Asp, Rossel, & Heim, 1996) as well as self-affirmation procedures (e.g., Fein & Spencer, 1997; Sherman, Kinias, Major, Kim & Prenovost, 2007; but see also Aberson, Healy, & Romero, 2000; Rubin & Hewstone, 1998) can improve intergroup attitudes.

In summary, evidence for the beneficial effects of the sophistication of social-cognitive skills exists, but is not as impressive as the findings concerning the impact of empathy evoking programs.

Besides the outcomes of interventions, results from survey-based research can give further important insights. As already mentioned, Pettigrew and Tropp (2008) conducted a meta-analysis of primarily survey-based studies to investigate the pathway from contact to reduced prejudice. Although, the objective of this research was to test mediators of contact-effects, the paths from knowledge to prejudice and empathy to prejudice (as well as from anxiety to prejudice) can be interpreted in a general way. The results clearly demonstrate that both knowledge and empathy predict the amount of prejudice with negative weights, however, the contribution of empathy (and of reduced anxiety) is much stronger.

To summarize, prior reviews do not provide satisfactory answers to the previously mentioned issues. However, based on the described theories with relevance to attitude change by means of information as well as on the presented findings, two assumptions can be made. First, programs that aim at the improvement of ethnic prejudice by means of information are generally effective. Second, initiatives that (also) address empathy can be expected to be more beneficial than those that do not include empathy-evoking content. The latter prediction is mainly based on the meta-analysis conducted by Pettigrew and Tropp (2008). Additionally, the consistency of findings concerning empathy programs is in line with it. These predictions, however, have not yet been confirmed and need to be tested with a contemporary and

comprehensive meta-analysis that systematically integrates the available evaluations of information interventions.

Methods of information programs. According to research reviewed by Oskamp (2000), methods that actively involve the participants typically lead to stronger and more persistent attitude changes. In accordance with that, some authors (Pedersen, Walker, Paradies, & Guerin, 2011; Pedersen, Walker & Wise, 2005; Stephan, 1999) argue that prejudice reduction programs with active elements more effectively improve ethnic attitudes than those that solely use passive procedures. Such an assumption is plausible, since the participants can—to some extent—contribute. Consequently, it can be expected that interventions with an active involvement of the participants are typically of higher personal relevance (Petty, Cacioppo, & Goldman, 1981) for them, thereby, enhance their motivation to elaborate the input and, in doing so, make a more beneficial attitude change possible. In line with that, the findings of qualitative research (e.g., Heppner & O'Brian, 1994) indicate that the participants of programs to reduce ethnic prejudice by means of active and passive strategies perceive the implemented active elements as especially effective. However, there is no review that provides a comprehensive test of the relative influence of various methods used to change ethnic attitudes (but see Stephan, Renfro, & Stephan, 2004).

Taken together, according to the available research it can be expected that interventions that (also) apply active procedures in order to improve ethnic attitudes are more effective than interventions that only use passive methods. Though, this prediction is yet to be confirmed.

Duration of information interventions. Finally, the issue of the influence of treatment length is of major relevance with regard to the optimal design of information programs as well as to the amount of resources that have to be invested. However, it can only be adequately addressed by reviewing the effectiveness of interventions that differ in

duration. In the light of these circumstances, it is surprising that there is no research synthesis that systematically addresses this question in the field of prejudice reduction programs (but see Stephan, Renfro, & Stephan, 2004). Though, based on the cursory inspection of some evaluations, authors of prior reviews (e.g., Pedersen, Walker, Paradies, & Guerin, 2011; Stephan, 1999; Stephan & Stephan, 1984) assume that longer interventions are more effective. Additionally, it was found (Malin, 1999) that participants of prejudice reduction strategies can have initial reservations. Furthermore, conceptually a more detailed treatment of the content of an intervention is possible when the participants are repeatedly exposed to it and the program is not a "one-shot" initiative or a very short initiative.

Summing up, the impact of the length of information interventions is yet to be clarified. Although there is no systematic evidence, it can be expected that longer information programs are typically associated with a better outcome.

The Present Meta-Analysis

Existing theories give reason to believe that information initiatives can be effective. Their actual impact, however, has to be scrutinized by means of outcome evaluations. Previous research does not provide a definitive answer regarding the general impact of information programs. In addition, although there is reason to assume that the consideration of empathy, the use of active elements, and a longer duration of the treatment are beneficial characteristics that can help to optimize the effect of information initiatives, available reviews on the reduction of ethnic prejudice can neither confirm nor disconfirm these predictions.

Considering the importance of these unresolved issues, we have conducted a comprehensive meta-analysis to test the following hypotheses:

Hypothesis 1: Information interventions generally reduce ethnic prejudice.

Hypothesis 2: Ethnic prejudice is more effectively reduced by information interventions that include an empathy-evoking component than by those that do not address empathy.

Hypothesis 3: Information programs which (also) utilize active methods have a better outcome than those that solely apply passive procedures.

Hypothesis 4: The duration of information interventions is positively associated with their impact.

To the best of our knowledge, the present work is the first meta-analytic review that tests these hypotheses by an exhaustive integration of the available literature on programs to reduce ethnic prejudice.

Method

Inclusion Criteria and Exclusion of Studies

We defined multiple criteria in order to specify the population of studies that are eligible for inclusion in the present meta-analytic integration. These criteria refer to five domains (independent variable; dependent variable; evaluation design; available data; language) and will be described in the following section. Since the aim of this research is to capture systematically the body of adequate studies on the outcome of information interventions, we refrained from determining further inclusion criteria regarding the age of the participants, the point of time the study was conducted, and the country in which the intervention was implemented.

1. Independent variable. To be considered, a study has to test the impact of an information intervention that was implemented in order to improve ethnic attitudes. As previously mentioned, the term information intervention refers to any program that provides some kind of input that is not based on intergroup contact. In accordance with that, we excluded structured contact programs (e.g., Schuitema & Veugelers, 2011; Slone, Tarrasch,

& Hallis, 2000) that have already been analyzed elsewhere (Lemmer & Wagner, 2011). Furthermore, we decided not to include studies (e.g., Bowman, 1977; McFadden, 1973) that insufficiently describe the content of the implemented program and therefore rendered it impossible to classify them at the content-axis.

2. Dependent variable. Relevant studies have to evaluate the outcome of an information intervention with at least one indicator of ethnic attitudes, that is, of attitudes toward members of an ethnic outgroup or toward an ethnic outgroup as a whole. Acceptable measures refer to the cognitive (e.g., beliefs), affective (e.g., sympathy), and/or behavior-oriented (e.g., readiness for contact) component of ethnic attitudes.

There are evaluations (e.g., Calandra, Fitzpatrick, & Barron, 2002; Thorman, 2003) that measured attitudes toward a specified ethnic group (e.g., African Americans) and report their results only for a (total) sample of participants that contains members of this group. We decided to include these studies when a clear majority of the participants (defined as at least 75%) do not belong to the group the dependent variable is concerned with. In addition, some evaluations (e.g., Rorrer & Furr, 2009; Walker-Dalhouse & Dalhouse, 2006) used measures that include both items with and without a clear interethnic reference. Regarding such configurations, we determined to include them when the majority of items (i.e., 75%) are clearly directed toward members of specific ethnic outgroups, toward specific ethnic outgroups as a whole, or towards ethnic outgroups in general (“other ethnic groups”). This criterion is not satisfied by some well-known instruments such as the “CDAI–Cultural Diversity Awareness Inventory (Henry, 1985)” and the total score of the “MEIM–Multigroup Ethnic Identity Scale (Phinney, 1992)”.

We did not consider studies that used self-developed measures and do not specify the content of them (e.g., Cascio & Bass, 1976; Rowe, 1976). In addition, we excluded evaluations that did not include at least one measure with a clear interethnic basis but solely

utilized dependent variables such as attitudes toward the own ethnic group or racial/ethnic identity (e.g., Berdugo, 1978; Brooks & Kahn, 1990). Furthermore, we did not incorporate studies (e.g., Amerson, 2010; Olson, Reed, & Schweinle, 2009; Sims, 1997) that merely used measures of self-assessed multicultural skills and knowledge such as the "MCI–Multicultural Counseling Inventory (Sodowsky, Taffe, Gutkin, & Wise, 1994, 1994)" or the "MCAS–Multicultural Counseling Awareness Scale (Ponterotto et al, 1996)". In our view, when used in the context of interventions, self-estimates of knowledge and skills with items such as "I am aware of my limitations in cross-cultural counselling and could specify them readily" are prone to substantial biases, for example, due to a potential motivation to reduce cognitive dissonance after investing in the participation in an intervention. Lastly, we did not include studies that are only based on qualitative dependent variables and do not report quantitative results together with evidence for an adequate interrater reliability for the quantification of the verbal material (e.g., McBride, 1997).

3. Evaluation design. We accepted studies with the following evaluation designs: *randomized posttest only with control (POWC)*, *pretest-posttest with control (PPWC)*, and *pretest-posttest single group (PPSG)*. The randomized POWC-design does not include a pretest but subjects are randomly assigned to the treatment or control group. By reason of randomization, discrepancies between the two groups are assumed to be eliminated except for one critical difference: treatment received vs. not received. Due to the existence of serious threats to internal validity (Shadish, Cook, & Campbell, 2002), we decided not to include studies (e.g., Weiner & Wright, 1973) that used a POWC-design without a random assignment of individuals[1] (i.e., at the level of analysis) to the intervention or to the control group. Moreover, we did not consider evaluations (e.g., Brisbin, 1971; Richeson & Nussbaum, 2003; Paluck, 2010) that applied a randomized posttest only design, but lacked an adequate control group. These studies compare different versions of an intervention and/or

include a “control” group that also received a—albeit weak—treatment with ethnic content. As the effect sizes originating from these evaluations do not model the actual net intervention effect and are therefore conceptually not comparable to the other studies, we decided to exclude them.

In addition to evaluations with a randomized POWC-design, we also included studies that utilized a randomized or non-randomized pretest-posttest with control design (PPWC). Since this design includes a pretest, initial differences between the intervention and the comparison group in the dependent variable can be controlled.

Furthermore, we considered studies with a pretest-posttest single group design (PPSG) that do not contain a control group. Although this design is—particularly because of potential history and maturation effects—marked by substantial threats to internal validity (Shadish, Cook, & Campbell, 2002), we decided not to exclude the corresponding studies from the beginning. Due to limiting factors, the utilization of a PPSG-design is often the only possibility to evaluate the impact of a social intervention. Therefore, such studies constitute a substantial portion of the research on information programs and should not be—in spite of their very problematic nature—ignored. Another reason for the accompanying inclusion of PPSG-comparisons is that this enabled us to accept studies that applied a pretest-posttest design with different treatment groups but no control group or with an intervention and a treated “control” condition (see above). In these cases, the pretest-posttest comparison for the actual intervention group was considered. Nonetheless, as will be described below, effect sizes that are based on PPSG-contrasts are analyzed separately from effect sizes that result from randomized POWC-designs and PPWC-designs.

4. Available data. We determined to solely consider studies that permit calculating an effect size with an appropriate degree of accuracy. Effect sizes of evaluation studies are calculated based on means, standard deviations, and sample sizes of the intervention and the

control group (see below). However, they can also be determined in an equivalent way from other basic statistics (e.g., t-test statistics) by using transformation formulae (e.g., Lipsey & Wilson, 2001). Nevertheless, in some cases an effect size cannot be calculated with an adequate degree of precision. Therefore, we were forced to exclude evaluations that only report results of complex multi-factorial ANOVA's or MANOVA's with repeated measurement (e.g., Breckheimer & Nelson, 1976; Elliot & Tyson, 1983; Black, 1973). In addition, in order to avoid biased representations, we did not consider studies (e.g., Bagley & Verma, 1972; Kraus, 1960; Randolph, Landis, & Tzeng, 1977; van der Keilen, 1977; Verma & Bagley, 1973; Verma & MacDonald, 1971; Weldon, Carlston, & Rissman, 1975; Weldon, Carlston, Rissman, Slobodin, & Triandis, 1974) that used multiple instruments to measure ethnic prejudice but did not allow the calculation of an effect size for all of them. Furthermore, we excluded studies with a PPSG-design that provide means (and standard deviations) for the pre- as well as for the posttest that are, however, based on differing sample sizes (e.g., Cole, 2003; Klein, 1992). Consequently, mean differences could not only be based on an actual change but also on a systematic attrition of participants.

5. Language. At last, we limited the population of relevant studies to documents that are written in either English or German.

Search for Relevant Literature

Our aim was to get as close as possible to the consideration of the entire population of studies that is defined by our inclusion criteria. Therefore, we utilized multiple strategies in order to find as many elements of that population as possible: searches in databases; manual searches in topic-related journals; consultation of organizations and experts; searches in conference proceedings; inspection of reference lists. The searches were performed repeatedly and research that was written up to the year 2010 was systematically covered.

1. Searches in databases. We queried international databases of multiple scientific disciplines: Psychology (e.g., PsycINFO, PSYNDEXplus), Education (e.g., ERIC - Education Resources Information Center, ERS - Education Research Complete), Social Sciences (e.g., Sociological Abstracts, SSCI - Social Sciences Citation Index, Social Services Abstracts), Media- and Communication Science (Communication & Mass Media Complete), Sports Science (SPORTDiscus), and Medicine (PubMed). In addition, we looked for documents in broad multidisciplinary international databases (e.g., Google Scholar, Scirus, WorldCat). Furthermore, since we aimed at finding as many unpublished documents as possible, we searched in multidisciplinary databases that include international as well as country-specific dissertations and master's theses (e.g., NLTD - Networked Digital Library of Theses and Dissertations, ProQuest Dissertations & Theses, Theses Canada, ULI - Israel Union Catalog, Australisian Digital Theses Program, Index to Theses, DATAD - Database of African Theses and Dissertations, DissOnline, DiVA - Academic Archive On-line, and NARCIS - National Academic Research and Collaborations Information System). Fourth, we queried specialized databases for grey literature (e.g., OpenSIGLE, NTIS - National Technical Information Service, NCJRS Abstracts Database, DTIC Public technical reports, CPCI - SSH - Conference Proceedings Citation Index, Social Sciences & Humanities).

Whenever possible, we utilized multiple structured and complex search algorithms in order to find entries that are eligible for inclusion. These algorithms contained terms that refer to essential elements[2] of relevant studies: (a) intervention component (e.g., intervent*, treatment*, training*, workshop*, program*, curriculum*), (b) ethnic component (e.g., ethnic*, racial*, cultural*, internat*), (c) dependent measure component (e.g., prejudic*, attitud*, stereotyp*, anxiet*, discriminat*, aggress*, violen*, relation*), and (d) evaluation component (e.g., evaluat*, effect*, outcome*, impact*, result*, reduc*, foster*, improv*). The usage of a multiplicity of synonyms within each component was necessary as there are

no standard terms that are typically mentioned in the titles or abstracts of appropriate studies—not even variations of the words “evaluation” and “intervention”. In order to find entries that contain in each block at least one of the synonyms, we combined the terms within the components with the boolean operator *or* and linked the blocks with the boolean operator *and*.

2. Manual searches in topic-related journals. In addition to the use of databases, we manually searched the content of journals that are associated with the topic of the present paper (e.g., Group Processes and Intergroup Relations, Intercultural Education, International Journal of Intercultural Relations, Journal of Applied Social Psychology, Journal of Educational Psychology). This accompanying search strategy was included in order to find further potential relevant literature that, for some reason, could not be detected with our search algorithms for the databases.

3. Consultation of organizations and experts. We used the listservs of a multiplicity of topic-related scientific organizations from multiple disciplines (e.g., Divisions of the American Psychological Association) to disseminate a call for sending us relevant research. Furthermore, we individually contacted scientific experts in the field of study from many countries. In addition, we consulted practitioners, evaluators, and organizations all over the world that are directly connected with interventions to improve interethnic relations.

4. Searches in conference proceedings. We also searched for appropriate research in the proceedings of topic-related conferences (e.g., ICPRI Peace education Conferences, ISPP annual meetings).

5. Inspection of reference lists. In addition to the described strategies, we manually checked the references of prior reviews that are, more or less, related to the topic under study (e.g., Aboud & Levy, 2000; Bigler, 1999; Denson, 2009; Engberg, 2004; Gudykunst, 1977, 1979, 1979; Lynch, 1985, 1987; McGregor, 1989; 1993; McGregor & Ungerleider, 1993;

Okoye-Johnson, 1999, Paluck & Green, 2009; Pedersen, Walker, Paradies, & Guerin, 2011; Pedersen, Walker, & Wise, 2005, Pendry, Driscoll, & Field, 2007; Ponterotto & Pedersen, 1993; Rose, 1948; Schofield, 1995; Stephan, 1999; Stephan, Renfro, & Stephan, 2004; Stephan & Stephan, 2001; St. Jean, 2007; Williams, 1947) as well as the reference list of each potentially relevant study that we found via the other search strategies.

Coding Procedure

The documents that were finally assessed as eligible for inclusion (see below) were subjected to a detailed coding process.

Study characteristics. In order to be able to provide a comprehensive description of the included research, to test our hypotheses two to four, and to investigate the influence of further study features, we coded formal characteristics, features of the interventions, as well as characteristics of the methodological quality and the dependent variables (see also Table 1).

1. Formal characteristics. With regard to formal features, we coded the year the document was written, the publication status of the document (journal article, book/book chapter, dissertation/master thesis, unpublished evaluation report), the country of the first author, and whether further documents exist that are directly related to the respective document.

2. Characteristics of the interventions. In order to systematically classify the type of program that was evaluated in the context of a given study, we used our multi-axial taxonomy. At the content axis, we coded for each study which of the nine sub-components (e.g., outgroup-focused knowledge; perspective taking; general cognitive skills) was addressed. This classification simultaneously determined at the superordinate level of content-components whether the intervention being evaluated in the respective document only dealt with one of the three components (i.e., solely knowledge-focused, empathy-

focused, or skill-focused), addressed two components (e.g., knowledge and empathy) or covered all three components. With reference to the method axis, we registered which passive (i.e., reading of texts and/or the inspection of visual material, listening to auditive information, viewing of audi-visual material, participating in lectures) and active (i.e., structured discussion, role play or simulation game, small group activity in order to prepare a "product", individual activity) procedures were utilized in the context of a given study to pass over the content. Consequently, the respective intervention was directly classified as one that applied only passive procedures, solely active techniques, or both passive and active methods (e.g., reading of texts and role plays). Furthermore, regarding the duration axis, we coded whether the respective program lasted up to one day or multiple days. In order to gain a more fine-grained—though not always ascertainable—index of the program length, we also calculated the number of net intervention hours as the product of the number of intervention days and the mean number of intervention hours on a given intervention day.

In addition to the program characteristics that refer to our multi-axial taxonomy, we classified the country in which the intervention was implemented and whether the program was realized by virtue of research interests or solely for practical reasons to solve social problems. Furthermore, we registered the age and sex (in % female) of the participants as well as their ethnic backgrounds.

3. Characteristics of the methodological quality and the dependent variables. Concerning the methodological quality, we, for instance, coded the type of the research design (i.e., POWC, PPWC, or PPSG), the level (i.e., individual or group) and the type (e.g., randomized, selected by others) of the assignment of participants to conditions, the nature of the control group (i.e., placebo treatment without ethnic reference, no special treatment or treatment as usual), the sample sizes, the extent of the attrition of the participants in the intervention as well as in the control group (i.e., % loss from the initial sample to the sample

that is considered in the data), whether a second posttest took place, the time intervals between the end of the intervention, the first posttest and a possible second posttest, as well as the total number of items that were used to measure ethnic attitudes. Furthermore, the measurement instruments used by the studies were classified considering three aspects: content, type of measure, and generalization. With regard to the content, we coded whether a given measure captures the cognitive component of ethnic attitudes, affective-behavioral aspects, or whether it represents a mixture of these categories and is therefore not clearly classifiable. We did not distinguish between the affective and behavior-oriented facet of attitudes. Our decision to combine these two aspects into a common category is based on the circumstance that these components could often not be adequately separated. In respect to the type of measure, we classified whether it is based on Likert attitude or social distance items, semantic differentials, used another type, or is mixed. Lastly, in reference to the level of generalization, we coded whether a given instrument focuses on one specific ethnic outgroup (e.g., African Americans), on diverse specific ethnic outgroups (e.g., African and Mexican Americans), on ethnic outgroups in general (e.g., "other ethnic groups"), or is mixed regarding the generalization level.

The coding process was structured by a detailed manual which contained a specification of the respective variables and categories, explicit coding rules, and typical examples. We determined the interrater reliability to test the quality of the developed coding system. Therefore, we trained a further coder who coded a random sample of about 15% of the included intervention-control comparisons. The interrater reliability was determined by using Cohen's κ (Cohen, 1960). This index corrects the agreement for the accordance expected by chance. The findings demonstrated that Cohen's κ for the considered study features is above 0.74 in every case, ranging from 0.75 (for "content of dependent variable") to 1.0 (e.g., "one-day vs. multi-day intervention", design). According to Landis and Koch

(1977) the determined interrater reliability can be classified as “substantial” to “almost perfect”.

Effect sizes. The effect sizes were calculated based on the information extracted from the texts or additionally provided by the author(s). As already described, we accepted three evaluation designs: randomized posttest only with control (POWC), pretest-posttest with control (PPWC), and pretest-posttest single group (PPSG). For that reason, we utilized three effect size indices to model the outcomes of the included studies: Hedges’s *g*, Morris’s *g*, and Becker’s *g*. These indices are bias-corrected modifications of the usual standardized mean difference, also known as Cohen’s *d*.

1. Hedges’s *g* (Hedges, 1981, 1982; Hedges & Olkin, 1985). In order to standardize the results from studies with a randomized POWC-design, we applied Hedges’s *g*. This index is defined as

$$g_i^{Hedges} = \left(\frac{(\bar{Y}_i^{post,I} - \bar{Y}_i^{post,C})}{\sqrt{\frac{(n_i^{post,I}-1)(s_i^{post,I})^2 + (n_i^{post,C}-1)(s_i^{post,I})^2}{n_i^{post,I} + n_i^{post,C} - 2}}} \right) \left(1 - \frac{3}{4(n_i^{post,I} + n_i^{post,C}) - 9} \right). \quad (1)$$

That is, for the *i*th study the posttest mean (denoted by $\bar{Y}$) of the control group (denoted by *C*) is initially subtracted from the posttest mean of the intervention group (denoted by *I*). This term is divided by the pooled posttest standard deviation of both groups. The result equals Cohen’s *d*. However, Hedges’s *g* multiplicatively adds a correction factor that prevents effects from studies with small samples from being upwardly biased and approximates “1” as sample sizes increase.

2. Morris’s *g* (Morris, 2008). Morris (2008) conducted a simulation study in order to compare alternative indices for the modeling of findings from studies with a PPWC-design. The following index has been shown to be the most beneficial and is labeled here as *Morris’s g*

$$g_i^{Morris} = \left(\frac{(\bar{Y}_i^{post,I} - \bar{Y}_i^{pre,I}) - (\bar{Y}_i^{post,C} - \bar{Y}_i^{pre,C})}{\sqrt{\frac{(n_i^{pre,I} - 1)(s_i^{pre,I})^2 + (n_i^{pre,C} - 1)(s_i^{pre,C})^2}{n_i^{pre,I} + n_i^{pre,C} - 2}}} \right) \left(1 - \frac{3}{4(n_i^{pre,I} + n_i^{pre,C} - 2) - 1} \right). \quad (2)$$

Morris's g for the *i*th study initially subtracts the pretest mean from the posttest mean (respectively denoted by $\bar{Y}$) both within the intervention group (denoted by *I*) and within the control group (denoted by *C*). The difference within the control group is then subtracted from the difference within the intervention group. The resulting term is divided by the pooled standard deviation at the time of the pretest and multiplied by a correction factor to avoid bias. Using Morris's *g* instead of Hedges's *g* to display the findings from evaluations with a PPWC-design has the advantage that pretest differences between the intervention and control group are eliminated which can be seen as the "standard of accuracy" (Carlson & Schmidt, 1999, p. 852).

3. Becker's *g* (Becker, 1988). We used Becker's *g* to standardize the results from studies with a PPSG-design. This index is given by

$$g_i^{Becker} = \left(\frac{\bar{Y}_i^{post,I} - \bar{Y}_i^{pre,I}}{s_i^{pre,I}} \right) \left(1 - \frac{3}{4(n_i^{pre,I} - 1) - 1} \right). \quad (3)$$

As can be seen, for the *i*th study it subtracts the pretest mean (denoted by $\bar{Y}$) of the intervention group (denoted by *I*) from the posttest mean of that group and divides the resulting term by the pretest standard deviation (denoted by *s*) of the intervention group. Again, a correction factor is multiplicatively added.

Each[3] of the three indices represents the difference between an intervention and a control condition in standard deviation units. A positive value indicates a positive effect of the respective information program (for reverse coded dependent variables the sign was changed).

Meta-Analytic Methods

Preliminary processing of effect sizes.

Clustering of effect sizes. We decided to meta-analyze the calculated effect sizes in two clusters: a *primary* and a *secondary cluster*. It can be assumed that the findings of evaluations that utilized a randomized POWC-design or a PPWC-design display the effect of the respective information programs in a way that is not affected by possible pretest discrepancies between the intervention and the control group as well as by possible time effects[4] (e.g., due to history and/or maturation). Consequently, Hedges's *g* and Morris's *g* can be assumed to estimate the intervention population effect without being biased by these aspects. For this reason, also legitimatized by research on meta-analytic methods (Morris & DeShon, 2002), we meta-analytically integrated the two designs and effect size indices, respectively, together in a primary cluster. In opposition to the results of evaluations that used a randomized POWC-design or a PPWC-design, the findings of studies that applied a PPSG-design can reflect both the intervention effect and a time effect. Therefore, the results of these studies and Becker's *g*, respectively, have to be regarded as potentially biased estimators of the intervention effect. Since they nonetheless constitute a substantial body of research on information programs, we analyzed them in a separate secondary cluster.

Unit of analysis. When it was possible, we determined detailed effects for different ethnic groups and age groups within the included studies. Calculating separate effect sizes for different ethnic groups within the studies enabled us to better investigate whether the impact of information programs varies for ethnic majority and ethnic minority samples. Computing separate effects for different age groups within the studies allowed us to better examine whether the outcome depends on the participant's age. Therefore, a given study can contribute with more than one effect size—each being based on a different sample of persons—to a meta-analytic integration. For this reason, we do not use study but

intervention-control comparison as the unit of our meta-analytic tests. This label refers to the contrast between an intervention group and a control condition, with the latter being an actual control group or a control baseline condition for studies with a PPSG-design.

Elimination of stochastic dependencies. Meta-analytic integrations require the absence of stochastic dependencies between the individual effect sizes (Hedges & Olkin, 1985; Matt & Cook, 2009). Consequently, a given sample of participants must contribute with just one effect size estimate to a given meta-analysis. We therefore used several procedures to compile a distribution of stochastic independent data points within the primary and within the secondary cluster.

1. In the event that a study reports findings concerning multiple relevant dependent variables for the same sample of participants, we computed an separate effect for each relevant measure and then aggregated the individual effects to obtain a single effect size for the particular sample (i.e., within the intervention-control comparison) prior to the meta-analytic integration. This method was also utilized in many other meta-analyses (e.g., Albarracín et al., 2005; Durantini, Albarracín, Mitchell, Earl, & Gillette, 2006; Mitte, 2005; Wilson & Lipsey, 2007; Wilson, Lipsey, & Derzon 2003). We preferred this procedure to alternative methods (e.g., randomly selecting one effect size per sample) since we assume that the aggregate of measures of ethnic attitudes is the best indicator of the effectiveness of the respective intervention.

2. Some evaluations compared different interventions to just one control group (e.g., Madden. 1970; Singh, 2004). The consideration of multiple effects in these cases—a separate effect size for the comparison of each intervention group with the same control group—would violate the assumption of stochastic independence[5] (Becker, 2000; Matt & Cook, 2009). In order to avoid biased estimates, we systematically selected[6] the intervention condition that—according to the hypothesis of the author(s)—was a priori expected to be the

most beneficial and therefore can be assumed to be the "optimal" information program in the respective study.

3. A few evaluations (e.g., Curl, 2002; Webster, 1994) have incorporated multiple control groups that received no treatment to improve interethnic relations (e.g., a first control group that received a placebo treatment and a second one that received no special treatment or was treated as usual) but just one intervention group. The simultaneous inclusion of the comparisons of each control sample with the same intervention sample would generate stochastic dependencies as just described. To conservatively test the impact of information programs, we systematically chose the control condition that is most similar to the intervention group, for example, the control group that received a placebo treatment (e.g., viewing a video that has no reference to the improvement of ethnic attitudes).

4. In order to eliminate dependencies that emerge when there is more than one posttest for a given intervention-control comparison, we separately meta-analyzed different time points of post measurement. In doing so, we considered two sets: *direct posttest* (less than one month after the end of the intervention) and *delayed posttest* (one month up to four[7] months after the end of the intervention). For some comparisons (e.g., Heath, 1992; Kimoto, 1974) data for multiple posttests within the direct posttest cluster is provided, in these cases we selected the test that was the closest to the end of the program.

Meta-analytic models. As stated in hypothesis 1, we expected a positive overall effect of information programs. This hypothesis was tested under the assumptions of the random effects model (REM; Hedges, 1983; Hedges & Olkin, 1985; Hedges & Vevea, 1998; Raudenbush, 1994, 2009; Viechtbauer, 2004). The REM suggests a two-level approach. At level one, the observed effect size of each individual study—or intervention-control comparison—estimates the study-specific true effect. The true effects of all relevant studies (i.e., that satisfy the inclusion criteria) are assumed to be randomly (normally) distributed at

level two with a mean that is the average true effect (denoted by μ_θ) and a variance that characterizes the variability (i.e., the heterogeneity) of the true effects (denoted by τ^2). The model equation at that level given by

$$\theta_i = \mu_\theta + \xi_i, \quad (4)$$

where θ_i represents the true effect of the *i*th study and μ_θ—as just mentioned—the average of the true effects of all studies. In addition, ξ_i denotes the discrepancy between the study-specific true effect and the average true effect. This term is represented by the variance of the true effects τ^2 when multiple studies are considered. As the true effects of the included studies are assumed to be a random sample drawn from the distribution of the true effects of all theoretically relevant studies and μ_θ is estimated as a parameter of that distribution, the REM allows unconditional inferences. That is, the results and conclusions are not restricted to the sample of the included studies but can be generalized to the hypothetical population of studies that satisfy the inclusion criteria. This population encompasses evaluations that are conducted and relevant but not included, could have been conducted and would be relevant, and that will be conducted in the future and will be relevant. In contrast, the competing fixed effects model (FEM; Hedges, 1983, 1994; Konstantopoulos & Hedges, 2009; Viechtbauer, 2004) holds that the included studies have a common true effect (denoted by θ) and that observed effect sizes differ solely because of their sampling errors. Therefore, no random effect variance exists and does not have to be considered in the meta-analytic integration. However, inferences deriving from the use of the FEM are restricted to the sample of the included studies. We have a priori decided to apply the REM since it allows the individual true effects to differ which is in our opinion more appropriate than the assumption of the FEM. Additionally, the REM permits a generalization of the findings.

Hypotheses 2 to 4 focus on the influence of intervention characteristics (e.g., empathy-evoking component included vs. not included) on the effectiveness of information programs. With reference to the consideration of potential moderators in a meta-analysis of research studies, two approaches exist: mixed effects model (MEM; Hedges & Olkin, 1985; Raudenbush, 1994, 2009; Viechtbauer, 2004) and fixed effects model with moderators (FEMwM; Hedges, 1994; Hedges & Olkin, 1985; Konstantopoulos & Hedges, 2009; Viechtbauer, 2004). A priori, we have determined to use the MEM. This model transfers the two-level approach of the REM to the fixed values of one (or more) potential moderator(s). Accordingly, within each value (e.g., empathy-evoking component included), the study-specific true effects are estimated by the corresponding observed effects and are randomly (normally) distributed around a value-specific average true effect size μ_θ (e.g., the average true effect for information programs that include an empathy-evoking component) with a variance τ^2 that represents the amount of heterogeneity that cannot be accounted for by the moderator(s). In accordance with our previous decision, we hold that the MEM is more appropriate than the FEMwM that transfers the assumptions of the FEM to the values of the potential moderator(s). In contrast to the FEMwM, the MEM also allows a generalization of the results to the population of (theoretically) relevant studies within the values of the potential moderator(s).

Meta-analytic procedures. We used a restricted maximum likelihood (REML) estimator of τ^2. Viechtbauer (2005) has demonstrated by means of Monte Carlo simulations that this estimator is efficient and has few biases. In order to examine the heterogeneity of the effects further, we applied Cochran's *Q*-Test for homogeneity (Cochran, 1954; Hedges & Olkin, 1985) as well as the I^2-statistic (Higgins & Thomas, 2002). A significant *Q*-test signals that the observed effect sizes differ more than is expected when sampling error would be the only variability-generating source. In these cases, it is typically concluded that heterogeneity

exists, that is, that the true effects differ. The I^2-statistic was introduced by Higgins and Thomas (2002) and models the amount of variability between the observed effects that is due to variability between underlying true effects, that is, due to heterogeneity.

Hypotheses 2 to 4 were tested with WLS meta-regression models (Steel & Kammeyer-Mueller, 2002; Thompson & Higgins, 2002; Viechtbauer, 2008). These models were specified under the assumptions of the MEM with dummy coded or continuous predictor variables.

In order to examine the robustness of the findings, we conducted sensitivity analyses. We tested the influence of the model choice by analyzing the data additionally under the rejected FEM and FEMwM. Furthermore, we examined the impact of potential outliers (i.e., effect sizes that are extreme relative to the others) by using externally standardized residuals (Hedges & Olkin, 1985, Viechtbauer, 2007, Viechtbauer & Cheung, 2010). An effect size with an absolute externally standardized residual that is larger than 1.96 can be regarded as a potential outlier (Viechtbauer & Cheung, 2010) that needs to be further investigated. Finally, we tested whether an over-estimation of the average true effect resulting from a potential publication bias (Rothstein, Sutton, & Borenstein, 2005) is likely. The best way to rule out such a threat to the validity of meta-analytic results is to conduct exhaustive searches for relevant literature. Since we have followed this principle, unpublished research is included to a substantial amount (see below). Nevertheless, we utilized methods to assess whether the results could be affected by a publication bias. First, we tested if the effects of published and unpublished documents differ systematically. In the event that no discrepancy exists, the typically assumed selective publication based on the desirability of the findings does not seem to be of a major concern in the content area of the present meta-analysis. Second, we inspected funnel plots (Light & Pillemer 1984; Light, Singer, & Willett, 1994) that plot the individual effect sizes against their corresponding standard errors. An asymmetric

distribution of the effect sizes around the mean true effect could signal that the distribution of included effect sizes is possibly biased. In this case, studies in the bottom left corner (i.e., studies that have lower effects than the average and have larger standard errors) are typically not included. Lastly, funnel plots were statistically tested for asymmetry with a rank correlation test[8] (Begg & Mazumdar, 1994) and a regression test[9] (Egger, Davey, Smith, Schneider, & Minder, 1997).

All analyses were conducted with the metafor package (Viechtbauer, 2010) for R (R Development Core Team, 2010).

Results

By means of the described search strategies, we identified 140 documents with 154 independent intervention-control comparisons (k = 106 within the primary cluster, k = 48 within the secondary cluster) that satisfy all criteria and that were therefore included in our meta-analytic integrations.

Table 1 displays central characteristics of the included research. Additional information is given in Appendix A (reference lists) and Appendix B (data concerning each intervention-control comparison of the primary and the secondary cluster).

General Effectiveness of Information Interventions

We predicted in hypothesis 1 that information interventions improve ethnic attitudes. In order to test this assumption, we separately assessed the general effectiveness of information programs within the primary cluster (i.e., with randomized POWC-designs and PPWC-designs) and within the secondary cluster (i.e., with PPSG-designs). In the primary cluster, we separately analyzed the impact at the time of the direct posttest (i.e., less than one month after the end of the intervention) and at the time of the delayed posttest (i.e., between one and four months after the end of the intervention). Since only one study of the secondary

cluster conducted a delayed posttest, this set was exclusively analyzed for the direct posttest. The findings can be seen in Table 2.

Results for the primary cluster.

Direct posttest. The meta-analytic integration is based on a total sample of $k = 103$ intervention-control comparisons with $N = 9{,}961$ participants (intervention: 5,584; control: 4,377). Figure 2 displays the distribution of the observed effect sizes under consideration of their standard errors. The average true effect is estimated to be $\hat{\mu}_\theta = 0.28$, 95% CI [0.22, 0.34]. The null hypothesis, stating that μ_θ is zero, can be rejected ($z = 9.43$, $p < .001$).

According to Cochran's *Q*-Test, the true effects of the individual comparisons differ ($Q = 274.26$, $df = 102$, $p < .001$). The variance of the true effects is estimated to be $\hat{\tau}^2 = 0.05$, 95% CI [0.03, 0.08]. The percentage of total variability between the observed effect sizes that is due to differences between the true effects is estimated to be $I^2 = 61.84\%$, 95% CI [49.39, 73.07] and can be characterized as "moderate to high" (Higgins & Thompson, 2002). Comparisons that are based on a randomized POWC-design ($\hat{\mu}_\theta = 0.28$, 95% CI [0.15, 0.41], $k = 25$) and those originating from a PPWC-design ($\hat{\mu}_\theta = 0.28$, 95% CI [0.22, 0.35], $k = 78$) do not differ ($Q_{model} = 0.001$, $df = 1$, $p = .98$) which empirically supports our joint analysis of these comparisons.

We conducted sensitivity analyses to examine the robustness of the reported results. First, we also tested hypothesis 1 under the assumptions of the Fixed Effects Model (FEM). The findings are almost identical to those obtained under application of the REM. The common true effect of the included studies is estimated to be $\hat{\theta} = 0.28$, 95% CI [0.25, 0.32] and differs significantly from zero ($z = 16.33$, $p < .001$). Second, we investigated the distribution of the individual effect sizes with regard to potential outliers. The absolute externally standardized residuals of four intervention-control comparisons (3.40, 2.47, 2.35,

2.22) exceed 1.96. Hence, these cases can be regarded as potential outliers. However, additional case deletion diagnostics demonstrated that meta-analyses without these comparisons do not result in substantially different estimates of μ_θ and heterogeneity. We therefore decided not to exclude the respective cases. Third, we examined with multiple procedures whether it is likely that the results are influenced by a publication bias. In doing so, we directly compared published comparisons ($\hat{\mu}_\theta = 0.28$, 95% CI [0.19, 0.37], $k = 44$) to unpublished comparisons ($\hat{\mu}_\theta = 0.29$, 95% CI [0.21, 0.36], $k = 59$). The two categories do not differ significantly ($Q_{model} = 0.01$, $df = 1$, $p = .93$). Furthermore, a funnel plot is displayed in Figure 2. An inspection of this plot for the primary cluster reveals that the observed effects are not asymmetrically dispersed around the average true effect. This visual impression is supported by statistical tests. Neither the rank correlation test (Kendall's $\tau = 0.08$, $p = .22$) nor the regression test ($z = 1.13$, $p = .26$) signal a funnel plot asymmetry.

Taken together, the presented findings for the primary cluster confirm hypothesis 1: Information interventions typically improve ethnic attitudes at the time of the direct posttest.

Delayed posttest. A sample of $k = 9$ intervention-control comparisons with a total of $N = 782$ participants (intervention: 372, control: 410) included a delayed posttest (i.e., one month up to four months after the end of the program). The estimated average true effect of these comparisons is $\hat{\mu}_\theta = 0.16$, 95% CI [0.03, 0.30] and is significantly different form zero ($z = 2.37$, $p < .05$).

Cochran's Q-Test signals an absence of heterogeneity ($Q = 8.43$, $df = 8$, $p = .39$). The estimated variance of the true effects is $\hat{\tau}^2 = 0.01$, 95% CI [0.00, 0.09]. In addition, an estimated "low" (Higgins & Thompson, 2002) percentage of $I^2 = 17.47\%$, 95% CI [0.00, 72.74] of the total variability between the observed effects is due to heterogeneity. Furthermore, the estimated average true effects of the two designs (i.e., randomized POWC-design and PPWC-design) do not differ ($Q_{model} = 0.96$, $df = 1$, $p = .33$).

Again, we conducted sensitivity analyses. An integration of the observed effects under the assumption of the FEM yielded an estimated true effect of $\hat{\theta} = 0.14$, 95% CI [0.03, 0.26] that is significant ($z = 2.38$, $p < .05$). Moreover, there are no potential outliers as the externally standardized residuals of all included comparisons are smaller than 1.96. Lastly, we examined whether the results are potentially influenced by a publication bias. Published comparisons ($\hat{\mu}_\theta = 0.26$, 95% CI [0.02, 0.50], $k = 3$) do not differ significantly ($Q_{model} = 1.05$, $df = 1$, $p = .31$) from unpublished comparisons ($\hat{\mu}_\theta = 0.14$, 95% CI [-0.03, 0.26], $k = 6$). The rank correlation test (Kendall's $\tau = 0.67$, $p < .05$) as well as the regression test ($z = 2.58$, $p < .05$) suggests that the funnel plot is asymmetric. However, an inspection of Figure 2 reveals that the overall pattern does not indicate a publication bias: Effect sizes in the lower left corner as well as in the upper right corner are missing.

In summary, the findings indicate a significant effect of intervention programs when measured more than one month after the end of the intervention.

Results for the secondary cluster.

Direct posttest. A total of $k = 48$ pretest-posttest comparisons including $N = 3{,}109$ participants were considered for the meta-analytic integration. The distribution of the observed effect sizes can be seen in Figure 2. The average true effect is estimated to be $\hat{\mu}_\theta = 0.22$, 95% CI [0.14, 0.29]. The null hypothesis, stating that μ_θ is zero, can be rejected ($z = 5.76$, $p < .001$).

According to Cochran's Q-Test, the true effects differ ($Q = 205.32$, $df = 47$, $p < .001$). The variance of the true effects is estimated to be $\hat{\tau}^2 = 0.05$, 95% CI [0.03, 0.11]. The estimated percentage of total variability between the observed effect sizes that is due to heterogeneity is $I^2 = 80.76\%$, 95% CI [73.97, 90.71] and can be categorized as "high" (Higgins & Thompson, 2002).

By the use of sensitivity analyses, we tested the robustness of the reported results. First, the data were also analyzed under the assumptions of the refused FEM. The estimate of the true effect is $\hat{\theta} = 0.14$, 95% CI [0.10, 0.17]. The null hypothesis can be rejected ($z = 9.14$, $p < .001$). Second, we inspected the distribution of the individual effect sizes with regard to potential outliers. Three of the included intervention-control comparisons have an absolute externally standardized residual (3.5, 2.32, 2.20) larger than 1.96 and can therefore be considered as potential outliers. Though, additional case deletion diagnostics showed that meta-analyses without these cases do not result in substantially different estimates of μ_θ and heterogeneity. The respective comparisons were hence not excluded. Lastly, we tested whether it is likely that the results are influenced by a publication bias. Therefore, we contrasted published ($\hat{\mu}_\theta = 0.23$, 95% CI [0.14, 0.33], $k = 29$) and unpublished ($\hat{\mu}_\theta = 0.19$, 95% CI [0.07, 0.32], $k = 19$) comparisons. No significant difference was found ($Q_{model} = 0.24$, $df = 1$, $p = .62$). The statistical tests of the shape of the funnel plot indicate an asymmetry (Kendall's $\tau = 0.23$, $p < .05$; regression test: $z = 3.14$, $p < .01$) which still exists after an exploratory deletion of the three potential outliers. However, a visual inspection of the funnel plot (see Figure 2) reveals that there is no severe distortion in the sense of a systematic absence of studies with smaller standard errors and effect sizes.

To summarize, we meta-analytically combined 48 effect sizes that are based on pretest-posttest comparisons without a control group. As for the primary cluster, the findings are in line with hypothesis 1: Information interventions have a positive impact on ethnic attitudes when determined less than one month after the end of the intervention.

Delayed posttest. As previously mentioned, a delayed posttest was conducted in the context of only one study. The observed effect of this study is $g = 0.25$, 95% CI [-0.04, 0.55] and differs marginally significant from zero ($z = 1.68$, $p < .1$).

Influence of Intervention Characteristics.

Test of Hypothesis 2 to 4. In addition to hypothesis 1, we predicted that the impact of information programs is influenced by essential program characteristics. More specifically, we hypothesized that the effectiveness of these interventions is higher when empathy is addressed (hypothesis 2) and participants are actively involved (Hypothesis 3). In addition, we predicted (hypothesis 4) that the duration of information initiatives is positively associated with their outcome.

Referring to our taxonomy, program length was operationalized with two indices: multi-day vs. one-day programs as well as the number of net intervention hours. Due to methodological shortcomings, the secondary cluster of effect sizes was not considered for the test of the predictions. By reason of the small number of comparisons with a delayed posttest, the analyses were also limited to the direct posttest.

The correlations between the four variables that are involved in our hypotheses are displayed in Table 3. It can be seen that although the use of active methods is conceptually independent from the duration of an intervention, the data show a strong empirical association: Interventions that (also) use active procedures are longer. Before we tested our hypothesis, we examined potential moderating influences of the study characteristics (see Table 1) that were coded but not involved in our hypotheses. None of these features did significantly influence the effect sizes. Therefore, no other variables were considered in the examination of our predictions.

We tested our hypotheses with mixed effects WLS meta-regression models, which have been shown with Monte Carlo simulations to be robust to even substantial multicollinearity and violations of the assumption of homoscedasticity (Steel & Kammeyer-Mueller, 2002). In this context, two sets of effect sizes were analyzed. The first set includes the total sample of 103 intervention-control comparisons of the primary cluster, the second

one consists of a sub-sample of 94 comparisons that allowed determining the number of net intervention hours.

We specified a multiple regression model (see Table 4) for the total sample and the three dummy coded potential moderators: empathy addressed (1 vs. 0), active methods used (1 vs. 0), and multi-day intervention (1) vs. one-day intervention (0).

The omnibus test of all potential moderators is significant ($Q_{model} = 45.50$, df = 3, $p < .001$, $k = 103$). The tests of the individual variables revealed a significant influence of empathy ($\hat{\beta} = 0.37$, $p < .001$), while the inclusion of active procedures ($\hat{\beta} = 0.04$, $p = .54$) and the usage of a multi-day design ($\hat{\beta} = -0.04$, $p = .59$) do not impact the effect sizes. In a further step, we included the two-way interaction terms between the three variables. The total model is significant ($Q_{model} = 46.70$, df = 6, $p < .001$, $k = 103$). Once again, empathy is the only significant moderator, while the interactions are insignificant.

We also fitted a multiple regression model for the reduced set of $k = 94$ effect sizes which enabled us to integrate the number of net intervention hours as an additional predictor. Again, the omnibus test is significant ($Q_{model} = 41.43$, df = 4, $p < .001$, $k = 94$). In regard to the individual predictors, the effect sizes increase when empathy is addressed ($\hat{\beta} = 0.38$, $p < .001$). In contrast, the contribution of the number of net intervention hours ($\hat{\beta} = 0.002$, $p = .26$) as well as of other two predictors is insignificant. A further analysis showed the absence of significant interactions.

In addition to these multiple regressions that respectively controlled for the influence of the other predictors and also investigated possible interactions, we conducted further univariate tests. Their results are displayed in Table 5 and contain the univariate sub-category based estimated average true effects that are typically given in the presentation of meta-analytic moderator analyses. Programs that include an empathy-associated component ($\hat{\mu}_0 =$

0.56, 95% CI [0.46, 0.65], $z = 11.65$, $p < .001$, $k = 27$) have a higher estimated average true effect ($Q_{model} = 45.78$, $df = 1$, $p < .001$) than interventions that do not address empathy ($\hat{\mu}_\theta =$ 0.18, 95% CI [0.12, 0.23], $z = 6.05$, $p < .001$, $k = 76$). Moreover, the impact of programs that apply active methods ($\hat{\mu}_\theta = 0.30$, 95% CI [0.23, 0.37], $z = 8.62$, $p = .05$, $k = 73$) and interventions that only use passive methods ($\hat{\mu}_\theta = 0.23$, 95% CI [0.11, 0.34], $z = 3.91$, $p <$ .001, $k = 30$) does not differ ($Q_{model} = 1.21$, $df = 1$, $p = .27$). Furthermore, there is no difference ($Q_{model} = 0.06$, $df = 1$, $p = .81$) between multi-day ($\hat{\mu}_\theta = 0.28$ (95% CI [0.21, 0.35], $z = 7.79$, $p < .001$, $k = 71$) and one-day ($\hat{\mu}_\theta = 0.29$ (95% CI [0.18, 0.40], $z = 5.26$, $p < .001$, k = 32) programs. Additionally, effect sizes are not influenced by the number of intervention hours ($Q_{model} = 0.25$, $df = 1$, $p = .61$, $k = 94$).

To sum up, the presented findings support hypothesis 2, that is, information interventions are more effective when they include an element that has empathy-evoking properties. This influence also exists when each of the other intervention features is statistically held constant in multiple mixed effects WLS meta-regressions. Moreover, since no interactions exist, the impact of empathy does not depend on whether the program includes the other intervention characteristics that were expected to be beneficial (i.e., active methods and prolonged treatment). In spite of a reduced effect, information interventions that do not focus on empathy improve ethnic attitudes, too. In contrast to hypothesis 2, the findings do not support hypothesis 3 and 4.

Additional analyses. In order to get further insights, we conducted additional analyses regarding the influence of the content, method, and duration of information interventions.

We tested whether the effect of the inclusion of an empathy-evoking component is restricted to dependent variables that include affective/behavioral content (0) or whether it is also evident for solely cognitive measures (1). The results of a moderated meta-regression

show a significant omnibus test (Q_{model} = 45.00, df = 1, p = .61, k = 103) and influence of empathy ($\hat{\beta}$ = 0.40, p < .001) but no significant interaction ($\hat{\beta}$ = -0.03, p = .77).

In addition, we analyzed the content of information strategies with more attention to detail. We compared the effectiveness of the three content configurations that were most often realized (see Table 1): solely knowledge-focused (reference category), knowledge- and skill-focused, as well as knowledge- and empathy-focused. The regression model (Q_{model} = 44.55, df = 2, p < .001) shows that programs that only focus on the enhancement of knowledge are effective ($\hat{\beta}$ = 0.16, p < .001), but contrasted to them, programs that address knowledge and skills ($\hat{\beta}$ = 0.19, p < .05) and interventions that include knowledge- and empathy-based content ($\hat{\beta}$ = 0.41, p < .001) are more effective. In addition, when contrasted, the latter configuration of components is more beneficial (p < .01) than initiatives that address knowledge and skills. The corresponding estimated average true effects are displayed in Table 5. Due to the multiplicity of possible sub-component configurations and the corresponding low numbers of cases, no further analyses were conducted.

We made additional investigations concerning our method axis. First, we eliminated those cases from the dataset that incorporate discussions as the only active technique. The reason for this is that discussions do not assure that most of the participants are really actively involved. The findings with these programs being excluded (Q_{model} = 1.05, df = 1, p = .31, k = 90) are similar to the results already presented. Second, with a more fine-grained analysis we compared the three configurations: solely passive (reference category), active and passive as well as active. Again, no significant difference was found as indicated by the omnibus test (Q_{model} = 2.11, df = 2, p = .35, k = 103). Third, we also tested, whether the number of different active and passive methods applied in a given study has an effect on the impact of information programs. The omnibus test of the regression model is insignificant (Q_{model} =

1.18, $df = 2$, $p = .55$, $k = 99$). Neither the number of different active strategies ($\hat{\beta} = 0.02$, $p = .36$) nor the number of passive techniques ($\hat{\beta} = .002$, $p = .96$) influences the effect sizes.

Concerning duration, the possibility exists that longer and shorter programs are comparably effective when analyzed immediately after the end of the intervention, but differ when tested with a certain temporal distance to the implementation of the program. This seems plausible since the effect of shorter interventions could soon diminish whereas longer programs could have a more persistent impact. In order to test whether such an interaction exists, we conducted two moderated meta-regressions. The first analysis was conducted with the total set and included three predictors: multi-day (1) vs. one-day (0) intervention, attitudes measured more than one day (1) vs. immediately (0) after the end of the program, and the interaction thereof. The model ($Q_{model} = 1.72$, $df = 3$, $p = .63$, $k = 103$) and the interaction are not significant ($\hat{\beta}_i = -0.05$, $p = .77$). The second meta-regression was conducted with the reduced set and considered net intervention hours instead of multi-day vs. one-day program. Again, the omnibus model test ($Q_{model} = 0.94$, $df = 3$, $p = .82$, $k = 94$) and the interaction term ($\hat{\beta}_i = -0.001$, $p = .80$) are not significant. Furthermore, we analyzed whether two other indices of duration (number of intervention days; gross intervention time) are associated with the effect sizes and also tested for non-linear associations. None of these analyses yielded significant findings.

To sum up, our results show—in addition to our test of hypothesis 2—an interesting sequence of the impact of information programs. Interventions that only address the expansion of knowledge are effective. However, the outcome of information interventions can be improved when a skill-focused component is added and even more when a component is added that has empathy-evoking properties. The latter configuration (empathy- and knowledge-focused) is the most beneficial. With further analyses we examined potential

contributions of the utilized methods and of treatment length, but these analyses failed to uncover additional effects.

Sensitivity analyses. The stated hypotheses were also tested under the assumptions of the FEMwM (fixed effects model with moderators). The resulting conclusions are in each case identical to those when analyzed in accordance with the MEM (mixed effects model).

Besides, we assessed the tested models with regard to potential outliers by using externally standardized residuals. There were only small numbers of potential outliers within the models. Analyses that were conducted without these cases led to results that do not substantially differ from those reported. Hence, we decided not to eliminate the respective cases.

Discussion

We used meta-analytic procedures to investigate the general effectiveness of information programs that were implemented in order to improve ethnic attitudes. In addition, we also examined the impact of central intervention characteristics.

General Effectiveness of Information Interventions

We expected (hypothesis 1) that programs that provide non-contact-based input in order to improve ethnic attitudes generally attain their goal. In order to test this prediction, we used effect sizes from 154 independent intervention-control comparisons. Among the considered effects, 106 effect sizes were based on studies with a randomized POWC-design or a PPWC-design and were thus assigned to a primary cluster. The other effect sizes originated from evaluations that only allowed comparing the posttest of the intervention group with its pretest and were analyzed in a separate secondary cluster.

Considering the direct posttest (i.e., less than one month after the end of the program), our results demonstrate that information interventions do in fact reduce prejudicial tendencies. In the primary cluster, an average true effect of $\hat{\mu}_{\theta} = 0.28$ was estimated. For the secondary

cluster, the estimated average effect is $\hat{\mu}_\theta = 0.22$. Both estimated mean true effects differ significantly from zero. To exemplify the effect in the primary cluster, an average participant of an information program is predicted to score on an ethnic attitude instrument with a standard deviation of two about 0.6 scale points (i.e., 28% of the standard deviation) better than an average control group member. With regard to the secondary cluster, the corresponding predicted difference is about 0.4 scale points (i.e., 22% of the standard deviation) in favour of an average intervention participant. Additionally, the estimated average true effects can be characterized as "small" to "medium" (Cohen, 1988) and as "educationally relevant" (Tallmadge, 1977).

While the primary cluster provides a clear and valid picture, the results obtained with the secondary cluster must be interpreted with caution. Since no control group is included, substantial threats to internal validity exist.

We also aimed to test whether the impact of information interventions is stable over time. In the primary cluster, nine intervention-control comparisons were meta-analytically integrated. An estimated mean true effect of $\hat{\mu}_\theta = 0.16$ resulted, which is significant, positive, and of "small" to "medium" size.

With regard to the secondary cluster, only one study included a delayed posttest, the point estimate is $g = 0.25$.

To summarize, our meta-analysis indicates that information programs that aim to reduce ethnic prejudice are effective when measured less than one month after the end of the intervention. In addition, the findings signal that the positive effect of information interventions is not restricted to a relatively short time period but is also observable when measured at least one month after the end of the intervention. However, no conclusions can yet be drawn regarding the persistence over longer time intervals since there are no studies

that included a delayed posttest that was conducted more than four months after the end of the program.

The Influence of Particular Intervention Components

We were also interested in testing whether particular components of information initiatives have an influence on their effectiveness. These components refer to the multi-axial taxonomy we propose. Given the problematic nature of the results obtained with the secondary cluster and the low number of delayed posttests, we restricted the corresponding analyses to the primary cluster and the direct posttest.

The importance of addressing empathy. Justified by previous research, we expected—as stated in the second hypothesis—that information programs that have empathy-evoking properties are more effective than interventions without them. The findings support our prediction. The estimated average true effect of initiatives with at least one empathy-associated component is significantly higher than the corresponding estimate of the average of the true effects of programs that are exclusively focused on knowledge- and/or skill-associated components. The difference is impressive and suggests the inclusion of empathy-evoking content in order to improve the effectiveness of information strategies. We also compared the impact of specific configurations. Due to low numbers of cases, some configurations (solely empathy-focused; solely skill-focused; empathy- and skill-focused; empathy-, skill- and knowledge-focused) could not be included. We found that interventions that address both empathy and knowledge have a higher impact than programs that focus on skills and knowledge which, in turn, are more effective than initiatives that only include knowledge-focused components. This finding suggests a clear pattern. Interventions that only address the expansion of knowledge do in fact improve ethnic attitudes, however, only to a modest degree. Focusing on empathy or skills does substantially increase the outcome of information programs. Our meta-analytic integration of the available research demonstrates

that the combination of empathy- and knowledge-based content is associated with the highest impact. Therefore, this configuration can be clearly recommended for future interventions.

Active methods do not improve the outcome. We assumed **(**hypothesis 3) that information programs that use at least one active procedure are more effective than interventions that only apply passive techniques. The results, though, do not confirm this prediction. Moreover, further analyses did also not support that the applied methods do systematically influence the impact of information initiatives. These findings are surprising as techniques with an active involvement of the participants can be expected to be especially motivating and therefore should enhance the chance of a beneficial outcome. Although research (e.g., Heppner & O'Brian, 1996) has shown that active elements are especially liked by participants and subjectively perceived as particularly effective by them, they do not improve the outcome of information interventions when tested meta-analytically with an integration of experimental and quasi-experimental evaluations. Even though active techniques do not significantly improve the impact of information programs, their incorporation can nonetheless be useful in order to generate a positive working atmosphere.

Duration does not matter. As stated in hypothesis 4, we expected that longer information interventions are more beneficial. In accordance with our multi-axial taxonomy, we used two different operationalizations of program length to test this hypothesis. First, we compared interventions that lasted one day at most with initiatives that were repetitively conducted on multiple days. The results demonstrate that one-day and multi-day programs have almost identical effects. Second, the influence of the number of net intervention hours as a further operationalization of duration was investigated. Again, no significant impact was found.

Even in the light of the presented results, the possibility of an interaction between program length and the time point of the posttest exists. With regard to this, an influence of

duration could emerge when the outcome of the intervention is measured with some temporal distance to its ending. The data, however, do not confirm this consideration.

The finding that duration doesn't matter is rather surprising and not easy to explain. In the face of time restrictions, shorter interventions are maybe characterized by very condensed and "catchy" elements. These aspects could compensate the lack of repetitive confrontations. Though, since it is not possible with the information given in the research documents to adequately determine the intensity of the implemented programs, this possibility could not be tested.

Taken together, our findings provide an unambiguous picture: content matters. Although information interventions are generally effective, the inclusion of a component with empathy-evoking properties substantially improves their impact. With regard to specific content configurations, we showed that interventions that incorporate both empathy- and knowledge-focused content are the most beneficial. On the other hand, the results do not confirm that the outcome of information programs is influenced by the type of the procedures used and program length. The presented findings are of great importance for policy and practice: Those programs should be implemented that address both the evocation of empathy and the expansion of knowledge. The effectiveness does not depend on the nature of the applied methods and on the duration of the programs. Note especially that the results concerning program length are advantageous in regard to the amount of resources and time that must be invested in the implementation of an effective information program.

Limitations

As with every research synthesis, this meta-analysis has some limitations.

First, it has to be considered that the studies included in the present review used self-reports to measure ethnic prejudice. Therefore, an obvious problem could be a social desirability bias (see Paulhus, 1984, 1991), that is, participants could have answered the

questions concerning ethnic attitudes in a way that corresponds more with a concern for being assessed favourably than with their actual attitudes. However, there is no reason why such a potential tendency should differently influence intervention and control groups. As a consequence, it has most likely not affected the observed effects this meta-analysis is based upon. The effect sizes that were applied by us model the difference between an intervention and a control condition and do not display the status of just one group. Another problem could be the influence of demand characteristics. Hence, the intervention participants could have noticed the hypothesis of the study (i.e., an improvement of ethnic attitudes by means of the implemented intervention) and could have (unconsciously) adjusted their answers at the posttest to it. Such a tendency would exclusively influence the means of the intervention conditions at the posttest and could therefore distort the calculated effect sizes. However, the obtained findings indicate that a substantial influence of demand characteristics is rather unlikely. Our results demonstrate—in line with our prediction—that programs with an empathy-evoking component have an impressively higher impact on ethnic attitudes than initiatives without such a component. There is no reason why especially empathy-evoking programs should be associated with stronger demand characteristics than programs that, for instance, explicitly deal with positive characteristics and accomplishments of ethnic outgroups.

Second, due to limitations in the primary research, we were not able to test mediations. That is, the degree in which the programs that are assumed to enhance knowledge, evoke empathy, and/or sophisticate social-cognitive skills really initiate these aspects and by reason of that reduce prejudicial tendencies.

Directions for Future Research

The present meta-analysis not only provides important findings that result from a systematic examination of the existing research on the effectiveness of information interventions but also gives impulses for future work on the impact of these programs.

First, as previously mentioned, no interventions are documented that explicitly deal with the reduction of intergroup anxiety and feelings of threat. However, survey-based research has demonstrated that these aspects are highly related to prejudicial tendencies. Therefore, ways to implement content that explicitly addresses anxiety and threat have to be explored.

Second, more research concerning the sustainability of the effects that are generated by information initiatives is needed.

Third, future research on the impact of information programs should systematically investigate mediating processes in order to be able to clarify the extent in which the different program types really work in the sense of the assumed mechanisms.

Fourth, it has to be examined if the absence of an effect of program length is due to the fact that the duration of information intervention has per se no influence or whether shorter interventions are as equally effective as longer interventions because they compensate a beneficial effect of program length by a higher intensity. We could not analyze his issue as intensity could not be determined with the descriptions provided in the documents. Future primary research could investigate this yet unanswered question by systematically manipulating the two factors—the length and intensity of an information intervention.

Conclusion

The results of this meta-analysis are of major importance for science as well as for policy and practice. We were able to demonstrate that information interventions do in fact improve ethnic attitudes. Furthermore, the impact of these programs is systematically

influenced by their content (axis one of our multi-axial taxonomy). In reference to this, information initiatives that include at least one component with empathy-evocating properties are more effective than programs without such an element. Furthermore, the combination of empathy- and knowledge-focused content has shown to be the most beneficial and can be recommended for future initiatives. Whereas the content of information interventions is significantly associated with their impact on ethnic prejudice, applied procedures (axis two) and length of treatment (axis three) are not.

References

Aberson, C. L., Healy, M., & Romero, V. (2000). Ingroup bias and self-esteem: A meta-analysis. *Personality and Social Psychology Review, 4*, 154–173.

Aboud, F. (1988). *Children and prejudice. London*: Blackwell Publishers.

Aboud, F. E., & Levy, S. R. (2000). Interventions to reduce prejudice and discrimination in children and adolescents. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 269–294). Mahwah, NJ: Erlbaum.

Albarracín, D., Gillette, J. C., Earl, A.N., Glasman, L. R., Durantine, M. R., & Ho, M. H. (2005). A test of major assumptions about behavior change: A comprehensive look at the effects of passive and active HIV–prevention interventions since the beginning of the epidemic. *Psychological Bulletin, 131*, 856–897.

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

Amerson, R. (2010). The impact of service-learning on cultural competence. *Nursing Education Perspectives, 31*, 18–22.

Ary, D. V., Tildesley, E., Hops, H., & Andrews, J. (1993). The influence of parent, sibling, and peer modeling and attitudes on adolescent use of alcohol. *International Journal of the Addictions, 28*, 853–880.

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). *Behaviour Research and Therapy, 33*, 825–831.

Bagley, C., & Verma, G. K. (1972). Some effects of teaching designed to promote understanding of racial issues in adolescence. *Journal of Moral Education, 1*, 231–238.

Bandura, A. (2002). Social cognitive theory of mass communication. In J. Bryant & D. Zillmann (Eds.), *Media effects: Advances in theory and research* (2nd ed., pp. 121–153). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Banks, J. A. (1984). Multicultural education and its critics: Britain and the United States. *The New Era, 65*, 58–64.

Banks, J. A. (1995). Multicultural education and the modification of students' racial attitudes. In W. D. Hawley & A. W. Jackson (Eds.), *Toward a Common Destiny: Improving Race and Ethnic Relations in America* (pp. 315–339). San Francisco, CA: Jossey-Bass.

Batson, C. D., & Ahmad, N. Y. (2009). Using empathy to improve intergroup attitudes and relations. *Social Issues and Policy Review, 3*, 141–177.

Batson, C. D., Chang, J., Orr, R., & Rowland, J. (2002). Empathy, attitudes, and action: Can feeling for a member of a stigmatized group motivate one to help the group? *Personality and Social Psychology Bulletin, 28*, 1656–1666.

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective taking: Imagining how another feels versus how you would feel. *Personality and Social Psychology Bulletin, 22*, 751–758.

Batson, C. D., Eklund, J. H., Chermok, V. L., Hoyt, J. L., & Ortiz, B. G. (2007). An additional antecedent of empathic concern: Valuing the welfare of the person in need. *Journal of Personality and Social Psychology, 93*, 65–74.

Batson, C. D., Polycarpou, M. P., Harmon-Jones, E., Imhoff, H. J., Mitchener, E. C., Bednar, L. L., Klein, T. R., & Highberger, L. (1997). Empathy and attitudes: Can feeling for a member of a stigmatized group improve feelings toward the group? *Journal of Personality and Social Psychology, 72*, 105–118.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257–278.

Becker, B. J. (2000). Multivariate meta-analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling*. (pp. 499–525). San Diego, CA: Academic Press.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101.

Berdugo, Y. (1978). *The design, implementation and evaluation of a theme-centered interaction program for adolescents in a Jewish day school* (Unpublished doctoral dissertation). University of Pittsburgh.

Bigler, R. S. (1999). The use of multicultural curricula and materials to counter racism in children. *Journal of Social Issues, 55*, 687–705.

Black, J. A. (1973). The effect of instruction in general semantics on ethnic prejudice. *Research in the Teaching of English, 7*, 98–108.

Black, J. S., & Mendenhall, M. (1990). Cross-cultural training effectiveness: A review and a theoretical framework for future research. *Academy of Management Review, 15*, 113–136.

Bowman, K. G. (1977). *Stereotypical responses of White undergraduate teacher candidates to Blacks before and after human relations training* (Unpublished doctoral dissertation). University of Minnesota.

Breckheimer, S. E., & Nelson, R. O. (1976). Group methods for reducing racial prejudice and discrimination. *Psychological Reports, 39*, 1259–1268.

Brewer, M. B., & Miller, N. (1984). Beyond the contact hypothesis: Theoretical perspectives on desegregation. In N. Miller & M. B. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 291–302). Orlando, FL: Academic Press.

Brisbin, C. D. (1971). *An experimental application of the galvanic skin response to the measurement of effects of literature on attitudes of fifth grade students toward Blacks* (Unpublished doctoral dissertation). Wayne State University.

Brooks, G. S., & Kahn, S. E. (1990). Evaluation of a course in gender and cultural issues. *Cultural Education and Supervision, 30*, 66–76.

Calandra, B., Fitzpatrick, J., & Barron, A. E. (2002). A Holocaust website: Effects on preservice teachers' factual knowledge and attitudes toward traditionally marginalized groups. *Journal of Technology and Teacher Education, 10*, 75-94.

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist, 22*, 817–829.

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of Applied Psychology, 84*, 851–862.

Cascio, W. F., & Bass, B. M. (1976). The effects of role playing in a program to modify attitudes toward Black employees. *Journal of Psychology: Interdisciplinary and Applied, 92*, 261–266.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*, 752–766.

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P Herman (Eds.), *Social influence: The Ontario symposium*, *Vol. 5* (pp. 3–39). Hillsdale, NJ: Erlbaum.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. *Unintended thought, 16*, 212–252.

Chapman, L. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior, 6*, 151–155.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101–129.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Cohen. J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cole, C. F, Arafat, C., Tidhar, C., & Tafesh, W. (2003). The educational impact of Rechov Sumsum/Shara'a Simsim: A Sesame Street television series to promote respect and understanding among children living in Israel, the West Bank and Gaza. *International Journal of Behavioral Development, 27*, 409–422.

Curl, L. S. (2002). *Can we reduce our latent prejudice? An examination of the Asian cultural assimilator with the use of the Implicit Association Test* (Unpublished doctoral dissertation). The University of Mississippi.

Dei, G. J. S. (1996). *Anti-racism education: Theory and practice*. Halifax, NS: Fernwood.

Denson, N. (2009). Do curricular and co-curricular diversity activities influence racial bias? A meta-analysis. *Review of Educational Research, 79*, 805–838.

Desai-Patel, S. P. (2005). *An evaluation of the interactive problem solving method* (Unpublished doctoral dissertation). Alliant International University, San Diego.

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. In J. F. Dovidio, M. Hewstone, P. Glick, & V. M. Esses (Eds.), *The Sage handbook of prejudice, stereotyping, and discrimination* (pp. 1–29). Thousand Oaks, CA: Sage.

Duckitt, J. (2010). Historical overview. In J. F. Dovidio, M. Hewstone, Glick, P., & V. M. Esses (Eds.), *The Sage handbook of prejudice, stereotyping, and discrimination* (pp. 29–45). Thousand Oaks, CA: Sage.

Durantini, M. R., Albarracín, D., Earl, A., & Mitchell, A. L. (2006). Conceptualizing the influence of social agents of behavior change: A meta-analysis of the effectiveness of HIV–prevention interventionists for different groups. *Psychological Bulletin, 132*, 212–248.

Eagly, A. H. (1987). *Sex differences in social behavior: A social–role interpretation*. Hillsdale, NJ: Erlbaum.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Fort Worth, TX: Harcourt, Brace, Jovanovich.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.

Elliot, G. A., & Tyson, G. A. (1983). The effects of modifying color-meaning concepts on the racial attitudes of Black and White South African preschool children. *The Journal of Psychology, 121*, 181–1990

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*, 473–524.

Esses, V. M., & Dovidio, J. F. (2002). The role of emotions in determining willingness to engage in intergroup contact. *Personality and Social Psychology Bulletin, 28*, 1202–1214.

European Union Agency for Fundamental Rights (2010). *Annual report on fundamental rights*. Retrieved from http://fra.europa.eu/fraWebsite/research/publications/

Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology, 73*, 31–44.

Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia, PA: Psychology Press.

Gaertner, S. L., Dovidio, J. F., & Houlette, M. A. (2010). Social categorization. In J. F. Dovidio, M. Hewstone, & V. M. Esses, (Eds.), *The Sage handbook of prejudice stereotyping and discrimination* (pp. 526–543). Thousand Oaks, CA: Sage.

Gaertner, S. L., Dovidio, J. F., Anastasio, P. A., Bachman, B. A., & Rust, M. C. (1993). The common ingroup identity model: Recategorization and the reduction of intergroup bias. *European Review of social Psychology, 4*, 1–26.

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57*, 239–249

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology, 78*, 708–724.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 339–355). New York, NY: The Russell Sage Foundation.

Gopaul-McNicol, S. (1986). *A cross-cultural study of the effects of modelling, reinforcerement and color meaning word association on doll color preference of Black preschool children and White preschool children in New York and Trinidad* (Unpublished doctoral dissertation). Hofstra University.

Govorun, O., Fuegen, K., & Payne, B. K. (2006). Stereotypes focus defensive projection. *Personality and Social Psychology Bulletin, 32*, 781–798

Grant, C. (1978). Education that is multicultural – Isn't that what we mean? *Journal of Teacher Education, 29*, 45–48.

Gudykunst, W. B. (1977). *The effects of an intercultural communication workshop on cross-cultural attitudes and interaction* (Unpublished doctoral dissertation). University of Minnesota.

Gudykunst, W. B. (1979). The effects of an intercultural communication workshop on cross-cultural attitudes and interaction. *Communication Education, 28*, 179–187.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392–407.

Heath, A. E: (1992). *The effects of training in racial identity stage theory on attitudes and perceptions of Whites toward Blacks* (Unpublished doctoral dissertation). University of Florida.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107–128.

Hedges, L. V. (1982). Estimation of effect size from a series of independent studies. *Psychological Bulletin, 92*, 490–499.

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93*, 388–395.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York, NY: The Russell Sage Foundation.

Hedges, L. V, & Olkin, I (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V, & Vevea, J. L (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.

Henry, G. (1986). *Cultural diversity awareness inventory=Invetorio sobre el reconocimiento de diversas culturas*. Unpublished manuscript.

Heppner, M. J., & O'Brien, K. M. (1994). Multicultural counselor training: Students' perceptions of helpful and hindering events. *Counselor Education and Supervision, 34*, 4–18.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539–1558.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin, 136*, 390–421.

Hogg, M.A., & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes*. London: Routledge.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communications and persuasion: Psychological studies in opinion change*. New Haven, CT: Yale University Press.

Iyer, A., & Leach, C. W. (2008). Emotion in inter-group relations. *European Review of Social Psychology, 19*, 86–125.

Katz, P. A. (1976). The acquisition of racial attitudes in children. In P. A. Katz (Ed.), *Towards the elimination of racism* (pp. 125–154). New York, NY: Pergamon.

Kimoto, C. K. (1974). *The effects of a juvenile literature based program on majority group attitudes toward Black Americans* (Unpublished doctoral dissertation). Washington State University.

Klauer, K. C., & Meiser, T. (2007). Training effects in deductive reasoning: A theory-based review. In W. Schaeken, A., Vandierendonck, W., Schroyens, & G. d'Ydewalle (Eds.),

The mental models theory of reasoning. Refinements and extensions (pp. 209–224). Mahwah, NJ: Lawrence Erlbaum.

Klein, T. E. (1982). *Teaching tolerance: Prejudice awareness and reduction in secondary schools* (Unpublished master's thesis). Dominican College (California).

Konstantopoulos, S., & Hedges, L. V. (2009). Analyzing effect sizes: Fixed effects models. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 279–295). New York, NY: The Russell Sage Foundation.

Kraus, S. (1960). Modifying prejudice: Attitude change as a function of the race of the communicator. *Audiovisual Communication Review, 10*, 14–22.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 1159–1174.

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The Communication of Ideas* (pp. 37–51). New York, NY: Harper & Row.

Lemmer, G., & Wagner, U. (2011). *Do direct and indirect contact programs improve ethnic attitudes? A meta-analytic evaluation of theory-driven interventions*. Manuscript submitted for publication.

Lewis, A. E. (2000). The impact of "colorblind" ideologies on students of color: Intergroup relations at predominantly white university. *The Journal of Negro Education, 69*, 74–91.

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Light, R. J., Singer, J. D., & Willett, J. B. (1994). Displaying and communicating findings from a meta-analysis. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 439–453). New York, NY: The Russell Sage Foundation.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (Applied Social Research Methods Series, Vol. 49). Thousand Oaks, CA: Sage.

Lynch, J. (1985). Human rights, racism, and the multicultural curriculum. *Educational Review, 37*, 141–152.

Lynch, J. (1987). *Prejudice reduction and the schools*. New York, NY: Nichols.

Mackie, D. M., Maitner, A. T., & Smith, E. R. (2009). Intergroup emotion theory. In T. D. Nelson (Ed.), *The handbook of prejudice, stereotyping, and discrimination* (pp. 285–308). New York, NY: Psychology Press.

Mackie, D. M., Smith, E. R., & Ray, D. G. (2008). Intergroup emotions and intergroup relations. *Personality and Social Psychology Compass, 2*, 1866–1880.

Madden, L. E. (1970). *Impact of information about Negroes on attitude change* (Unpublished doctoral dissertation). Ball State University.

Malin, M. (1999). I'm rather tired of hearing about it . . .': Challenges in constructing an effective anti-racism teacher education program. *Curriculum Perspectives, 19*, 1–11.

Martin, I. & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy, 1*, 57–102.

Martin, I., & Levey, A. B. (1985). Conditioning, evaluations and cognitions: An axis of integration. *Behaviour Research and Therapy, 23*, 167–175.

Martin, I., & Levey, A. B. (1987). Learning what will happen next: Conditioning, evaluation and cognitive processes. In G. Davey (Ed.), *Cognitive processes and Pavlovian conditioning in humans* (pp. 57–82). New York, NY: John Wiley.

Masi, C. M., Chen, H.-Y., Hawkley, L. C., & Cacioppo, J. T. (in press). A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review*.

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 537–560). New York, NY: The Russell Sage Foundation.

McBride, R. A. (1997). *Multicultural education utilizing multiethnic children's literature in fourth grade classrooms: A descriptive case study* (Unpublished doctoral dissertation). Oklahoma State University, Stillwater, OK.

McFadden, J. (1973). A*n investigation to determine the effects of two group training methods on racial prejudice* (Unpublished doctoral dissertation). University of South Carolina.

McGregor, J. (1989). *Strategies to reduce racial prejudice in students: A meta-analysis of research* (Unpublished doctoral dissertation). McGill University (Canada).

McGregor, J. (1993). Effectiveness of role playing and antiracist teaching in reducing student prejudice. *Journal of Educational Research, 86*, 215–226.

McGregor, J., & Ungerleider, C. (1993). Multicultural and racism awareness programs for teachers: A meta-analysis of the research. In K. A. McLeod (Ed.), *Multicultural education: The state of the art national study report #1* (pp. 59–63). Toronto, ON: University of Toronto, Faculty of Education.

Meiser, T., & Hewstone, M. (2010). Contingency learning and stereotype formation: Illusory and spurious correlations revisited. *European Review of Social Psychology, 21*, 285–331.

Mendenhall, M. E., Stahl, G. K., Ehnert, I., Oddou, G., Osland, J. S., & Kühlmann, T. M. (2004). Evaluation Studies of Cross–Cultural Training Programs. A Review of the Literature from 1988 to 2000. In D. Landis, J. M. Bennett, & M. J. Bennett (Eds.), *Handbook of intercultural training* (pp. 129–144). Thousand Oaks, CA: Sage.

Miller, N. (2002). Personalization and the promise of contact theory. *Journal of Social Issues, 58,* 387–410.

Mitte, K. (2005). Meta-analysis of cognitive behavioral treatments for generalized anxiety disorder: A comparison with pharmacotherapy. *Psychological Bulletin, 131*, 785–795.

Morris, S. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364–386.

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

Okoye-Johnson, O. N. (1999*). A meta-analysis of the effects of multicultural education on the racial attitudes of pre–k through grade 12 students: A comparison of curricular intervention and reinforcement dimensions in suburban and urban educational settings* (Unpublished doctoral dissertation). Old Dominion University.

Olson, S., Reed, K., & Schweinle, A. (2009). Improving multicultural awareness. *Journal of Inquiry and Action in Education, 2*, 120–139.

Oskamp, S. (2000). Interventions to reduce prejudice and discrimination in children and adolescents. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 269–294). Mahwah, NJ: Erlbaum.

Oskamp, S. (2000). *Attitudes and opinions (2nd ed.)*. Englewood Cliffs, NJ: Prentice Hall.

Oskamp, S., & Schultz, P. W. (2005). *Attitudes and Opinions (3nd ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Paluck, E. L. (2010). Is it better not to talk? Group polarization, extended contact, and perspective taking in eastern Democratic Republic of Congo. *Personality and Social Psychology Bulletin, 36*, 1170–1185.

Paluck, E. L., & Green, D. P. (2009). Prejudice Reduction: What works? A critical look at evidence from the field and the laboratory. *Annual Review of Psychology, 60*, 339–367.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598–609.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press, Inc.

Pate, G. S. (1981). Research on prejudice reduction. *Educational Leadership, 38*, 288–291.

Pedersen, A., Walker, I., & Wise, M. (2005). Talk does not cook rice: Beyond anti-racism rhetoric to strategies for social action. *Australian Psychologist, 40*, 20–30.

Pedersen, A., Walker, I, Paradies, Y., & Guerin, B. (2011). How to cook rice: A review of ingredients for teaching anti-prejudice. *Australian Psychologist, 46*, 55–63.

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.

Pettigrew, T. F., & Tropp, L. R. (2008). How does intergroup contact reduce prejudice? Meta-analytic tests of three mediators. *European Journal of Social Psychology, 38*, 922–934.

Pettigrew, T. F., & Tropp, L .R. (2011). *When groups meet: The dynamics of intergroup contact*. London, UK: Routledge/Psychology Press.

Petty, R. E., & Cacioppo, J. T. (1986a). *Communication and persuasion: Central and peripheral routes of attitude change*. New York, NY: Springer.

Petty, R. E., & Cacioppo, J. T. (1986b). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, *Vol. 19*. Orlando, FL: Academic Press.

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. D. Gilbert, S. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., pp. 323–390). New York: McGraw–Hill.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determiant of argument-based persuasion. *Journal of Personality and Social Psychology, 41*, 847–855.

Phinney, J. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with adolescents and young adults from diverse groups. *Journal of Adolescent Research, 7*, 156–176.

Piaget, J., & Weil, A. M. (1951). The development in children of the idea of the homeland and of relations to other countries. *International Social Science Journal, 3*, 561–578.

Ponterotto, J. G., & Pedersen, P. B. (1993). *Preventing prejudice: A guide for counselors and educators*. Newbury Park, CA: Sage.

Ponterotto, J. G., Rieger, B. P., Barrett, A., Sparks, R., Sanchez, C. M., & Magids, D. (1996). Development and initial validation of the Multicultural Counseling Awareness Scale. In G. R. Sodowsky & J. C. Impara (Eds.), *Multicultural assessment in counseling and clinical psychology* (pp. 247–282). Lincoln, NE: Buros Institute of Mental Measurements.

Prakash, A. O. (1972). *The effect of modeling and social reinforcement on the racial preferences of children* (Unpublished doctoral dissertation). University of Montana.

R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria.

Randolph, G., Landis, D., & Tzeng, O. (1977). The effects of time and practice on culture assimilator training. *International Journal of Intercultural Relations, 1*, 105–119.

Raudenbush, S. W. (1994). Analyzing effect sizes: Random effects models. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 302–321). New York, NY: The Russell Sage Foundation.

Raudenbush, S. W. (2009). Analyzing effect sizes: Random effects models. In H. Cooper, L. V. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–316). New York, NY: The Russell Sage Foundation.

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology, 40*, 417–423.

Rorrer, A. S., & Furr, S. (2009). Using film as a multicultural awareness tool in teacher Education. *Multicultural Perspectives, 11*, 162-168.

Rose, A. M. (1948). *Studies in reduction of prejudice: A memorandum summarizing research on modification of attitudes*. Chicago, IL: American Council on Race Relations.

Rosenthal, R., & Rubin, D. B., (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*, 400–406.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.

Rowe, F. A. (1976). *A study of the effects of a race relations education program on attitudes of racial prejudice* (Unpublished doctoral dissertation). University of Southern California.

Rubin, M., & Hewstone, M. (1998). Social identity theory's self-esteem hypothesis: A review and some suggestions for clarification. *Personality and Social Psychology Review, 2*, 40–62.

Schaller, M. (1994). The role of statistical reasoning in the formation, preservation, and prevention of group stereotypes. *British Journal of Social Psychology, 33*, 47–61.

Schaller, M., & O'Brien, M. (1992). "Intuitive analysis of covariance" and group stereotype formation. *Personality Social Psychology Bulletin, 18*, 776–785.

Schaller, M., Asp, C. H., Rosell, M. C., & Heim, S. J. (1996). Training in statistical reasoning inhibits the formation of erroneous group stereotypes. *Personality and Social Psychology Bulletin, 22*, 829–844.

Schofield, J. W. (1986). Causes and consequences of the color-blind perspective. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination and racism* (pp. 231–253). New York, NY: Academic Press.

Schofield, J. W. (1995). Improving intergroup relations among students. In J. A. Banks & C. A. M. Banks (Eds.), *Handbook on multicultural education* (pp. 628–634). New York, NY: Macmillan Publishing.

Schuitema, J., & Veugelers, W. (2011). Multicultural contacts in education: a case study of an exchange project between different ethnic groups. *Educational Studies, 37*, 101–114.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton–Mifflin.

Sherman, D. K., Kinias, Z., Major, B., Kim, H. S., & Prenovost, M. (2007). The group as a resource: Reducing biased attributions for group success and failure via group affirmation. *Personality and Social Psychology Bulletin, 33*, 1100–1112.

Sherman, D., & Cohen, G. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology, 38*, 183–242.

Shih, M., Wang, E., Trahan Bucher, A., & Stotzer, R. (2009). Perspective taking: Reducing prejudice towards general outgroups and specific individuals. *Group Processes Intergroup Relations, 12*, 565–577.

Short, G., & Carrington, B. (1996). Anti-racist education, multiculturalism, and the new racism. *Educational Review, 48*, 65–77.

Sims, J. M. (1981). *An exploratory study of the effects of human relations training on the attitudes toward Blacks of White teacher educators as measured by the Multifactor Racial Attitude Inventory* (Unpublished doctoral dissertation). The University of Iowa.

Singh, S. (2004). *The role of personal relevance on racial prejudice reduction: An application of the elaboration likelihood model* (Unpublished doctoral dissertation). The University Of Akron.

Skinner, B.F. (1953). *Science and Human Behavior*. New York, NY: Macmillan.

Skinner, B. F. (1972). *Cumulative record: a selection of papers*. New York, NY: Appleton-Century-Crofts.

Slone, M., Tarrasch, R., & Hallis, D. (2000). Ethnic stereotypic attitudes among Israeli children: Two intervention programs. *Merrill-Palmer Quarterly, 46*, 370–389.

Sodowsky, G. R., Taffe, R. C., Gutkin, T. B., & Wise, S. L. (1994). Development of the Multicultural Counseling Inventory: A self-report measure of multicultural competencies. *Journal of Counseling Psychology, 41*, 137–148.

St. Jean, G. (2007). *Evaluation of intervention programs designed to address high school racial conflicts* (Unpublished doctoral dissertation). George Mason University.

Steel, P. D., & Kammeyer-Mueller, J. D. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96–111.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261–302.

Stephan, C. W., Renfro, L., & Stephan, W. G. (2004). The evaluation of multicultural education programs: Techniques and a meta-analysis. In W. G. Stephan, & W. P. Vogt (Eds.), *Education programs for improving intergroup relations: Theory, research, and practice* (pp. 227–242). New York, NY: Teachers College Press.

Stephan, W. G. (1999). *Reducing prejudice and stereotyping in schools* (Multicultural Education Series). New York, NY: Teachers College Press.

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

Stephan, W. G., & Stephan, C. W. (1984). The role of ignorance in intergroup relations. In N. Miller & M. B. Brewer (Eds.), *Groups in contact: The psychology of desegregation* (pp. 229–257). Orlando, FL: Academic Press.

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–45). Mahwah, NJ: Erlbaum.

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

Stephan, W. G., Stephan, C. W., & Gudykunst, W. B. (1999). Anxiety in intergroup relations: A comparison of anxiety/uncertainty management theory and integrated threat theory. *International Journal of Intercultural Relations, 23*, 613–628.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Monterey, CA: Brooks/Cole.

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

Tallmadge, G. K. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education, and the US Office of Education.

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21*, 1559–1573.

Thorman, A. E. (2002). *An evaluation of the impact of anti-bias curriculum practices on prejudicial racial attitudes among children attending Head Start* (Unpublished doctoral dissertation). Kansas State University.

Tropp, L. R., & Pettigrew, T. F. (2004). Intergroup contact and the central role of affect in intergroup prejudice. In C. W. Leach & L. Tiedens (Eds.), *Social life of emotion* (pp. 246–269). Cambridge, UK: Cambridge University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D. & Wetherell, M. S. (1987). *Rediscovering the social group*. Oxford, UK: Blackwell.

U.S. Department of Justice (2009). *Hate crime statistics of the Federal Bureau of Investigation*. Retrieved from http://www2.fbi.gov/ucr/hc2009/index.html

U.S. Department of State (2009). *The country reports on human rights practices*. Retrieved from http://www.state.gov/g/drl/rls/hrrpt/2009/index.htm

Van der Keilen, M. (1977). Some effects of a special Indian culture oriented program on attitudes of White and Indian elementary school pupils. *Canadian Journal of Behavioural Science, 9*, 161–168.

Verma, G. K., & Bagley, C. (1973). Changing racial attitudes in adolescents: An experimental English study. *International Journal of Psychology, 8*, 55–58.

Verma, G. K., & MacDonald, B. (1971). Teaching race in schools: Some effects on the attitudinal and sociometric patterns of adolescents. *Race, 13*, 187–202.

Vescio, T. K., Sechrist, G. B., & Paolucci, M. P. (2003). Perspective taking and prejudice reduction: the mediational role of empathy arousal and situational attributions. *European Journal of Social Psychology, 33*, 455–472.

Viechtbauer, W. (2004). *Choosing between the fixed-, random-, and mixed-effects model in meta-analysis: An analysis of existing and new model selection methods* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics, 30*, 261–293.

Viechtbauer, W. (2007, May). *Outlier detection in meta-analysis via studentized deleted residuals*. Paper presented at the Annual Campbell Colloquium, London, England.

Viechtbauer, W. (2008). Analysis of moderator effects in meta-analysis. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 471–487). Thousand Oaks, CA: Sage.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods, 1*, 112–125.

Wagner, U., & Hewstone, M. (in press). Intergroup contact. In L. R. Tropp (Ed.), *The Oxford handbook of intergroup conflict*. New York, NY: Oxford University Press.

Wagner, U., Christ, O., & van Dick, R. (2002). Die empirische Evaluation von Präventionsprogrammen gegen Fremdenfeindlichkeit [The empirical evaluation of primary prevention programs of hostility against foreigners]. *Journal für Konflikt– und Gewaltforschung, 4*, 101–117.

Walker, H. (1989). Towards anti-racist, multicultural practice with underfives. *Early Child Development and Care, 41*, 103–112.

Walker-Dalhouse, D., & Dalhouse, A. D. (2006). Investigating White preservice teachers beliefs about teaching in culturally diverse classrooms. *Negro Educational Review, 57*, 69–784.

Webster, C. B. (1994). *Antiracism training through the exploration of White racial identity* (Unpublished doctoral dissertation). University of Georgia

Weiner, M. J., & Wright, F. E. (1973). Effects of undergoing arbitrary discrimination upon subsequent attitudes toward a minority group. *Journal of Applied Psychology, 3*, 94–102.

Weldon, D. E., Carston, D. E., Rissman, A. K., Slobodin, L., & Triandis, H. C. (1974). *A laboratory test of the effects of culture assimilator training*. Unpublished Manuscript.

Weldon, D. E., Carston, D. E., Rissman, A. K., Slobodin, L., & Triandis, H. C. (1975). A laboratory test of effects of culture assimilator training. *Journal of Personality and Social Psychology, 32*, 300–310.

Wilder, D. A. (1978). Reducing intergroup discrimination through individuation of the outgroup. *Journal of Personality and Social Psychology, 36*, 1361–1374.

Whitely, B. E., & Kite, M. E. (2006). *The psychology of prejudice and discrimination*. Belmont, CA: Thomson Wadsworth.

Williams, R. M., Jr. (1947). *The reduction of intergroup tensions*. New York, NY: Social Science Research Council.

Wilson, S. J., & Lipsey, M. J. (2007). Effectiveness of school-based intervention programs on aggressive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine 33(Suppl. 2)*, 130–143.

Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology, 71*, 136–149.

Footnotes

[1] Since in the field of evaluations the number of groups is usually small, a randomization at the group-level cannot be seen as an adequate alternative.

[2] We decided not to use target words that are connected with the design of the evaluation studies because many documents do not contain explicit statements concerning the applied evaluation design in their title and abstract.

[3] In contrast to Hedges's *g*, the calculation of a standard error for Morris's *g* and Becker's *g* technically requires the pre-post correlation. Since pre-post correlations were rarely reported, we decided in accordance with our procedure in a further meta-analysis (Lemmer & Wagner, 2011) to set the correlation to .7. This value approximates the average test-retest correlations of attitude scales reported in the literature, is also used in other meta-analyses (e.g., Masi, Chen, Hawkley & Cacioppo, 2010).

[4] The absence of an interaction between the factors "intervention vs. control" and "time" is assumed.

[5] Dependency problems do not arise when a study includes more than one treatment group and a separate control group for each treatment group. In this case, all intervention-control comparisons can be included in the meta-analytic integration as they are based on different samples.

[6] Because different interventions within one study are most often based on different conceptions, we selected one intervention and did not use an aggregate across intervention conditions.

[7] For none of the included comparisons, a delayed posttest with a time interval of more than four months after the end of the intervention was conducted.

[8] This test is based on the rank correlation between the standardized effect sizes and the transformed standard errors. A significant rank correlation (Kendall's τ) indicates an

association between the two variables. In case of asymmetry, high standard errors should be systematically associated higher effect sizes.

[9] The method provided by Egger et al. (1997) utilizes the inverse of the standard error to predict an index that is calculated as the effect size divided by the corresponding standard error. In case the intercept differs significantly from zero, results could be biased.

Appendix A

References of the Documents Considered in the Meta-Analysis Integrations

A1

References for the Primary Cluster

Aboud, F. E, & Fenwick, V. (1999). Exploring and evaluating school-based interventions to reduce prejudice. *Journal of Social Issues, 55*, 767–785.

Aguilar, J. V. (1973). *A comparative study of teaching methods for developing more positive attitudes in elementary school children toward ethnic minority groups* (Unpublished doctoral dissertation). The University of Nebraska-Lincoln.

Albey, R. H. (1986). *The effects of interracial communication training on Anglo supervisors' job performance ratings of minority employees* (Unpublished doctoral dissertation). University of California, Los Angeles.

Alsbrook, E. Y. (1970). *Changes in the ethnocentrism of a select group of college students as a function of bibliotherapy* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Aoki, J. (1991*). Effects of the culture assimilator on cross-cultural understanding and attitudes of college students* (Unpublished doctoral dissertation). Southern Illinois University at Carbondale.

Backa, E. A. (1997). *The effects of exposure to anti-bias stories on the racial attitudes of preschool-age children* (Unpublished doctoral dissertation). University of South Florida.

Bazelak, L. P. (1973). *A content analysis of tenth-grade students' responses to Black literature, including the effects of reading this literature on attitudes towards race* (Unpublished doctoral dissertation). Syracuse University.

Bennett, C. T. (1979). The preparation of pre-service secondary social studies teachers in multiethnic education. *High School Journal, 62*, 232–237.

Bickley, A. R. (1974). *A study of the effects of teaching a unit on Black culture to classes of predominantly White high school students* (Unpublished doctoral dissertation). West Virginia University.

Bird, A., Heinbuch, S., Dunbar, R., & McNulty, M. (1993). A conceptual model of the effects of area studies training programs and a preliminary investigation. *International Journal of Intercultural Relations, 17*, 415–435.

Blanchard, F. A., Crandell, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology, 79*, 993–997.

Blanchard, F. A., Lilly, T., & Vaughn, L. A. (1991). Reducing the expression of racial prejudice. *Psychological Science, 2*, 101–105.

Bolton, E. B. (1935). Effect of knowledge upon attitudes towards the Negro. *Journal of Social Psychology, 6*, 68–90.

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

Cadd, M. (1994). An attempt to reduce ethnocentrism in the foreign language classroom. *Foreign Language Annals, 27*, 143–160.

Cernat, V. (2001). I know what you did last millennium: Ethnic stereotype and attitude change after reminding people of historical events. *Current Research in Social Psychology, 7*, No Pagination Specified.

Curl, L. S. (2002). *Can we reduce our latent prejudice? An examination of the Asian cultural assimilator with the use of the Implicit Association Test* (Unpublished doctoral dissertation). The University of Mississippi.

Cushner, K. H. (1989). Assessing the impact of a culture-general assimilator. *International Journal of Intercultural Relations. 13*, 125–146.

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology, 90*, 784–803.

Davidson, J. R. (1991). *Evaluation of an education model for race/ethnic sensitive social work and critique of the White Racial Identity Attitude Scale* (Unpublished doctoral dissertation). The University of Texas at Arlington.

Davis, E. E., & Fine, M. (1975). The effects of the findings of the U.S. national advisory commission on civil disorder: An experimental study of attitude change. *Human Relations, 28*, 209–227.

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

Duck, J. M., Lalonde, R., & Weiss, D. (2003). International images and mass media: The effects of media coverage on Canadians' perceptions of ethnic relations in Australia. *Australian Journal of Psychology, 55*, 15–23.

Earnest, C. W. (1968). *The effects of two teaching methods in modifying ninth-grade pupil's attitudes toward minority groups* (Unpublished doctoral dissertation). The University of Oklahoma.

Esses, V. M., & Dovidio, J. F. (2002). The role of emotions in determining willingness to engage in intergroup contact. *Personality and Social Psychology Bulletin, 28*, 1202–1214.

Feinman, G. (1982). *The impact of humanistic education instruction on self-concept, racial prejudice, school attendance, school performance, and discipline* (Unpublished doctoral dissertation). Fordham University.

Feuchte, F. (2010). "*Can we make a difference tomorrow?" A systematic evaluation of a peace education programme implemented with Liberian refugees* (Unpublished doctoral dissertation). University of Jena, Germany.

Geißler, N. (2008). *Konzeption und Evaluation eines Unterrichtsprogramms zur Prävention und Verminderung von Antisemitismus* (Unpublished doctoral dissertation). University of Potsdam, Germany.

Goerke, F. D. (1984). *Schüler und Ausländerfeindlichkeit. Vorurteilsreduktion bei Hauptschülern durch sozialgeographishe Unterrichtsprojekte* (Unpublished doctoral dissertation). University of Bonn, Germany.

Graves, S. B. (1975). *Racial diversity in children's television: Its impact on racial attitudes and stated program prefererences in young children* (Unpublished doctoral dissertation). Harvard University.

Gray, D. B., & Ashmore, R. D. (1975). Comparing the effects of informational, role-playing, and value-discrepant treatments on racial attitudes. *Journal of Applied Social Psychology, 5*, 262–281.

Greenberg, H., Person, J., & Sherman, S. (1957). The effects of single-session education techniques on prejudice attitudes. *Journal of Educational Sociology, 31*, 82–86.

Gwinn, C. B. (1998). *The effect of multicultural literature on the attitudes of second grade students* (Unpublished doctoral dissertation). University of Minnesota,

Hamarnehm, M. E. (1995). *The impact of racial identity training on anti-black attitudes of White counselors-in-training* (Unpublished doctoral dissertation). University of Florida.

Hawkins, J., & Frechtling, J. (1981). *Short-term effects of human relations training (HR-17): A pre post comparison group evaluation study*. Unpublished Manuscript.

Heath, A. E. (1992). *The effects of training in racial identity stage theory on attitudes and perceptions of Whites toward Blacks* (Unpublished doctoral dissertation). University of Florida.

Hines, B. M. (1984). *A study of the effect of children's literature with nonstereotyped Black characters on the racial attitudes of sixth-grade children* (Unpublished doctoral dissertation). Northwestern State University of Louisiana.

Hoffmann, C., & Scheffler, D. (2005). *Evaluation der Wirkung und Zielerreichung von Seminaren der BildungsBausteine gegen Antisemitismus*. Unpublished Manuscript.

Hohn, R. L. (1973). Perceptual training and its effects on racial preferences in kindergarten children. *Psychological Reports, 32*, 435–441.

Holland, R. F. (1969). *The effects of two study units on high school students' attitudes toward Negroes* (Unpublished doctoral dissertation). The University of Oklahoma.

Houser, B. B. (1978). An examination of the use of audiovisual media in reducing prejudice. *Psychology in the Schools, 15*, 116-122.

Jones, J. B. (1983). *The identification of changes in racial attitudes of Black and White students as the result of the utilization of specific curriculum experiences* (Unpublished doctoral dissertation). State University of New York at Buffalo,

Karafantis, D. M. (2004). *The impact of colorblind principles on the social attitudes of college students* (Unpublished doctoral dissertation). State University of New York at Stony Brook.

Keats, B. E. (1989). *The effects of an Afro-American history course on the racial attitudes of high school students* (Unpublished doctoral dissertation). The University of Wisconsin-Madison.

Kehoe, J. W., & Rogers, W. T. (1978). The effects of principle testing discussions on student attitudes toward selected groups subjected to discrimination. *Canadian Journal of Education, 3*, 73–80.

Kimoto, C. K. (1974). *The effects of a juvenile literature based program on majority group attitudes toward Black Americans* (Unpublished doctoral dissertation). Washington State University.

Klein, E. B. (1963). Stylistic components of response as related to attitude change. *Journal of Personality*, 31, 38–51.

Klitzke, A. (2005). *Evaluation eines Antidiskriminierungstrainings: Ziele und Wirkungen des Blue Eyed Workshops aus sozialpsychologischer Sicht* (Unpublished master's thesis). University of Jena, Germany.

Koeller, S. (1977). The effects of listening to excerpts from children's stories about Mexican-Americans on the attitudes of sixth graders. *Journal of Educational Research, 70*, 329–334.

Kraemer, A. J. (1975). *Vicarious attitude change and the design of 'message' films: Application to race relations*. Unpublished Manuscript.

Laghaie, R. F. (1991). *The effects of factual information on the attitudes of people toward a given culture: An American and Iranian example* (Unpublished doctoral dissertation). Ball State University

Lemmer, G., & Wagner, U. (2003). *Evaluation des Präventionsprogrammes "Cool und taff"*. Unpublished Manuscript.

Levinson, D. J. (1954). The intergroup relations workshop: Its psychological aims and effects. *The Journal of Psychology, 38*, 103–126.

Levy, S. R., West, T. L., Bigler, R. S., Karafantis, D. M., Ramirez, L., & Velilla, E. (2005). Messages about the uniqueness and similarities of people: Impact on U.S. Black and Latino youth. *Journal of Applied Developmental Psychology*, *26*, 714–733.

Litcher, J. H., & Johnson, D. W. (1973). Use of pictures of multiethnic interaction to change attitudes of White elementary school students toward Blacks. *Psychological Reports, 33*, 367–372.

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

Madden, L. E. (1970). *Impact of information about Negroes on attitude change* (Unpublished doctoral dissertation). Ball State University.

Maluso, D. (1992). *Interventions to lessen racist prejudice and discrimination among college students* (Unpublished doctoral dissertation). University of Rhode Island.

Maxson, R. (1970). *The effect of a structured learning experience on the attitudes of selected high school students toward minority ethnic groups* (Unpublished doctoral dissertation). Mississippi State University

May, J. J. (1997). *The impact of a multicultural-racial awareness intervention on selected preservice teachers' cultural perceptions of African-Americans and other selected traditionally underrepresented populations* (Unpublished doctoral dissertation). Bowling Green State University.

McIlveen, Y. C. L. (1988). *An assessment of the ability of the culture-general assimilator to create sensitivity to multiculturalism in an educational setting* (Unpublished doctoral dissertation). University of Houston.

Megumi, S. (1994). *A comparison of the effects of multicultural and anti-racist teaching on student attitude change* (Unpublished doctoral dissertation). University of British Columbia, Vancouver (Canada).

Neto, F. (2006). Changing intercultural attitudes over time. *Journal of Intercultural Communication, 12*, No Pagination Specified.

Newell, K. R. (2001). *The effects of multicultural literature on elementary students' social distance attitudes toward selected racial groups* (Unpublished doctoral dissertation). The University of Southern Mississippi.

Okyne, E. S. (1980). *Teaching African Studies in American high schools* (Unpublished doctoral dissertation). Union Institute and University.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*, 421–433.

Orson, C. M. (1971). *Effects on self-concept and racial prejudice of a structured intervention program with desegregated fifth grade children* (Unpublished doctoral dissertation). University of Miami.

Page, E. R. (1974). *An analysis of racial atitude change in eleventh grade students using a black literature package* (Unpublished doctoral dissertation). The Pennsylvania State University.

Parish, T. S. (1974). Conditioning of racial attitudes and color concepts in children. *Perceptual and Motor Skills, 39*, 707–714.

Patruno, M. (2001). *The effectiveness of multicultural storytelling on attitudes toward different cultures in the elementary classroom* (Unpublished doctoral dissertation). Rowan University.

Peacock, B. D. (1992). *The effects of an ethnic music program on attitudes of racial acceptance* (Unpublished doctoral dissertation). University of Massachusetts Lowell.

Pearl, D. (1955). Psychotherapy and ethnocentrism. *Journal of Abnormal and Social Psychology, 50*, 227–229.

Persson, A., & Musher-Eizenman, D. R. (2003). The impact of a prejudice-prevention television program on young children's ideas about race. *Early Childhood Research Quaterly, 18*, 530–546.

Pettijohn, T. F., & Walzer, A. S. (2009). Reducing racism, sexism, and homophobia in college students by completing a psychology of prejudice course. *The Open Social Science Journal, 2*, 1–6.

Probst, T. M. (2003). Changing attitudes over time: Assessing the effectiveness of a workplace diversity course. *Teaching of Psychology, 30*, 236–239.

Schaefer, C. (2004). *Kooperatives Lernen als Methode zur Reduzierung von Vorurteilen am Beispiel des Antisemitismus – Eine Studie zur Validierung eines Interventionsprogramms* (Unpublished master's thesis). University of Potsdam, Germany.

Schiele, D. T. (1979). *The impact of a specific educational experience on the modification of White toward Black racial attitudes* (Unpublished doctoral dissertation). University of Denver.

Schrüfer, G. (2003). *Verständnis für fremde Kulturen* (Unpublished master's thesis). University of Bayreuth, Germany.

Schwartz, C. S. L. (1972).*The effect of selected black poetry on expressed attitudes toward Blacks of fifth-grade and sixth-grade White suburban children* (Unpublished doctoral dissertation). Wayne State University.

Shirley, O. L. B. (1988). *The impact of multicultural education on the self-concept, racial attitude, and student achievement of Black and White fifth and sixth graders* (Unpublished doctoral dissertation). The University of Mississippi.

Shiver, W. B. (1970). *The effectiveness of two forms of an experimental teaching unit in reducing black ninth and eleventh graders' negative interracial attitudes toward white people* (Unpublished doctoral dissertation). The University of Oklahoma.

Singh, S. (2004). *The role of personal relevance on racial prejudice reduction: An application of the elaboration likelihood model* (Unpublished doctoral dissertation). The University of Akron.

Smey, J. W. (1983). *A model to improve racial attitudes and relationships of racial prejudice with behavior during a health care intervention* (Unpublished doctoral dissertation). Clark University.

Soble, J. R., Spanierman, L. B., & Liao, H. L. (2011). Effects of a brief video intervention on White university students' racial attitudes. *Journal of Counseling Psychology,* 58, 151–157.

Sousa, M. D. R., Neto, F., & Mullet, E. (2005). Can music change ethnic attitudes among children? *Psychology of Music, 33*, 304–316.

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.

Tauran, R. H. (1967). *The influences of reading on the attitudes of third graders toward Eskimos* (Unpublished doctoral dissertation). University of Maryland College Park.

Timmermans, B. B. (2000). *Development of cultural sensitivity in undergraduate nursing students* (Unpublished doctoral dissertation). The University of New Mexico.

Turner, R. N., & Brown, R. (2008). Improving children's attitudes toward refugees: An evaluation of a school-based multicultural curriculum and an anti-racist intervention. *Journal of Applied Social Psychology, 38*, 1295–1328.

Vrij, A., Akehurst, L., & Smith, B. (2003). Reducing ethnic prejudice: An evaluation of seven recommended principles for incorporation in public campaigns. *Journal of Community and Applied Social Psychology, 13*, 284–299.

Wagner, U., & Homburg, A. (2004). *Abschlussbericht zur Evaluation des Xenos-Projektes "Horizonte Erweitern" von Arbeit und Bildung e.V. Marburg/Lahn und Arbeit und Bildung Nordhausen e.V*. Unpublished Manuscript.

Walker, P. A. (1971). *The effects of hearing selected children's stories that portray Blacks in a favorable manner on the racial attitudes of groups of Black and White kindergarten children* (Unpublished doctoral dissertation). University of Kentucky.

Webster, C. B. (1994). *Antiracism training through the exploration of White racial identity* (Unpublished doctoral dissertation). University of Georgia.

West, M. A. (1997). *Long-term effects of a prejudice-reduction simulation on attitude change* (Unpublished doctoral dissertation). Long Island University, The Brooklyn Center.

Wilke, R. L. (1997). *A multicultural curriculum's relationship to attitudes of prejudice and stereotyping in 5th and 6th-grade students* (Unpublished doctoral dissertation). University of Southern California.

Williams, D. M. (1945). *A study of the relative effectiveness of selected teaching procedures in the modification of children's attitudes toward the Negro* (Unpublished doctoral dissertation). New York University.

Williams, H. M. (1961). Changes in pupils' attitudes towards West African Negroes, following the use of two different teaching methods. *British Journal of Educational Psychology. 31*, 292–296.

Yawkey, T. D. (1973). Attitudes toward Black Americans held by rural and urban White early childhood subjects based upon multi-ethnic social studies materials. *Journal of Negro Education, 42*, 164–169.

A2

References for the Secondary Cluster

Aoki, D. A. M. (2000). *The impact of a multicultural counseling training course on students' multicultural identity development levels* (Unpublished doctoral dissertation). University of Denver.

Ayres, M. (1973). Counteracting racial stereotypes in preschool children. *Graduate Research in Education and Related Disciplines, 6*, 55–74.

Baba, Y., & Hebert, C. (2004). The effects of participation in a cultural awareness program on Jail Inmates. *Journal of Ethnic and Cultural Diversity in Social Work, 13*, 91–113.

Boatright-Horowitz, S. L. (2005). Teaching antiracism in a large introductory psychology Class: A course module and its evaluation. *Journal of Black Studies, 36*, 34–51.

Broaddus, D. L. (1986). *Use of the culture-general assimilator in intercultural training* (Unpublished doctoral dissertation). Indiana State University.

Brown, S. P., Parham, T. A., & Yonker, R. (1996). Influence of cross-cultural training course on racial identity attitudes of White women and men: Preliminary perspectives. *Journal of Counseling & Development, 74*, 510–516.

Butler, D. L. (1982). *The reduction of racial prejudice through the use of vicarious experience* (Unpublished doctoral dissertation). University of Michigan.

Case, K. A. (2007). Raising White privilege awareness and reducing racial prejudice: Assessing diversity course effectiveness. *Teaching of Psychology, 34*, 231–235.

Clinton, L. T. (1983). *The effectiveness of two methods of teaching sociology in facilitating students' rational thinking about racial and ethnic groups* (Unpublished doctoral dissertation). East Texas State University.

Cullen, J. E. (2008). "*Some friends and I started talking…" - A participatory action research project to deconstruct white privilege among student affairs practitioners* (Unpublished doctoral dissertation). Iowa State University.

Darigan, D. L. (1991). *The effects of teachers reading aloud on elementary children's attitudes toward African Americans* (Unpublished doctoral dissertation). University of Oregon.

Dine, D. B. (1994). *Prejudice reduction with grade primary and one students: A comparison of multicultural and anti-racist teaching strategies* (Unpublished doctoral dissertation). Saint Mary's University (Canada).

Droba, D. D. (1932). Education and Negro attitudes. *Sociology and Social Research, 17*, 137–141.

Elley, W. B. (1964). Attitude change and education for international understanding. *Sociology of Education, 37*, 318–325.

Fife, P. (1994). *The effects of American Indian cultural awareness training on the stereotyping behaviors of preservice teachers* (Unpublished doctoral dissertation). University of Arkansas.

Gordon, E. R. (1983). *Television's potential for reducing racial prejudice in children: A cognitive moral developmental approach* (Unpublished doctoral dissertation). The Claremont Graduate University.

Hayes, M. T. (1969). *An investigation of the impact of reading on attitudes of racial prejudice* (Unpublished doctoral dissertation). Boston University School of Education.

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 and Applied Social Psychology, 11*, 243–262.

Hussey, H. D., Fleck, B. K. B., & Warner, R. M. (2010). Reducing student prejudice in diversity-infused core psychology classes. *College Teaching, 58*, 85–92.

Johnson, M. R. (1980). *An exploration of preservice teacher perceptions on the effectiveness of a multicultural lecture series* (Unpublished doctoral dissertation). University of Michigan.

Korhonen, K. (2002). *Intercultural competence as part of professional qualifications: A training experiment with Bachelor of Engineering students*. (Unpublished doctoral dissertation). University of Jyväskylä, Finnland.

Kuperus, C. G. (1992). *Using interaction with children's literature to positively affect the global and multicultural/multiethnic knowledge and attitudes of third-grade students*. Unpublished Manuscript.

Lefley, H. P. (1985). Impact of cross-cultural training on Black and White mental health professionals. *International Journal of Intercultural Relations, 9*, 305–318.

Levinson, D. J., & Schemerhorn, R. A. (1951). Emotional-attitudinal effects of an intergroup relations workshop on its members. *Journal of Psychology 31*, 243–256.

Lillis, J., & Hayes, S. C. (2007). Applying acceptance, mindfulness, and values to the reduction of prejudice: A pilot study. *Behavior Modification, 31*, 389–411.

Lopez, G., Gurin, P., & Nagda, B. R. (1998). Education and understanding structural causes of group inequalities. *Political Psychology, 19*, 305-329.

Martin, P. S. (2001). *The effect of antiracism education on the racial identity development and White privilege awareness in White teachers* (Unpublished doctoral dissertation). University of Massachusetts Boston.

McMahon, R., & Reeves-Kazelskis, C. (1995). *Effects of children's literature on preservice teachers' opinions about multicultural education issues*. Unpublished Manuscript.

Neville, H. A., Heppner, M. J., Louie, C. E., & Thompson, C. E. (1996). The impact of multicultural training on White racial identity attitudes and therapy competencies. *Professional Psychology: Research and Practice, 27*, 83–89.

Pedersen, A., Aly, A., Hartley, L., & McGarty, C. (2009). An intervention to increase positive attitudes and address misconceptions about Australian Muslims: A call for education and open mindedness. *The Australian Community Psychologist, 21***,** 81–93.

Pedersen, A., & Barlow, F. (2008). Theory to social action: A university based strategy targeting prejudice against Aboriginal Australians. *The Australian Psychologist, 43*, 148–159.

Sanchez, T. R. (1991). *The effects of knowledge acquisition about Blacks on the racial attitudes of White high school sophmores*. (Unpublished doctoral dissertation). Indiana University.

Singh, J. M., & Yancey, A. V. (1974). Racial attitudes in White first grade children. *Journal of Educational Research, 67*, 370–372.

Smith, M. (1943). A second report on changes in attitudes toward the Negro. *School and Society, 57*, 388–392.

Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychological Bulletin, 27*, 484–494.

Steed, R. (2010). Attitudes and beliefs of occupational therapists participating in a cultural competency workshop. *Occupational Therapy International, 17*, 142–151.

Stenson, S. A. (1978). *The effect of a unit in Chicano literature on ninth-grade students' attitudes toward Chicanos*. (Unpublished doctoral dissertation). The University of Tennessee.

Tuttle, H. G., Guitart, J., Papalia, A., & Zampogna, J. (1979). Effects of cultural presentations on attitudes of foreign language students. *The Modern Language Journal, 63*, 177–182.

Van Soest, D. (1996). The impact of social work education on student attitudes and behavior concerning oppression. *Journal of Social Work Education, 32*, 191–202.

Wieder, G. (1954). Group procedures modifying attitudes of prejudice in the college classroom. *Journal of Educational Psychology, 45*, 332–344.

Wittig, M. A., & Molina, L. (2000). Moderators and mediators of prejudice reduction. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 295–318). Mahwah, NJ: Erlbaum.

Appendix B

Characteristics of the Included Intervention-Control Comparisons

Table B2

Characteristics of the Intervention-Control Comparisons of the Primary Cluster

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
1.	Aboud	1999	Pub.	K, S	A	M	11	PPWC	—	—	80	0.317
2.	Aguilar	1973	Unpub.	K	A, P	M	—	PPWC	80	-0.063	—	—
3.	Albey	1986	Unpub.	K, S	A, P	M	—	PPWC	17	0.189	—	—
4.	Alsbrook	1970	Unpub.	K, E	A, P	M	37.5	PPWC	26	1.287	—	—
5.	Aoki	1991	Unpub.	K	A, P	O	1	POWC	41	0.362	—	—
6.	Backa	1997	Unpub.	K	A, P	M	20	PPWC	33	0.003	—	—
7.	Bazelak	1973	Unpub.	K, E	A, P	M	8	PPWC	130	-0.172	—	—
8.	Bennett	1979	Pub.	K	A, P	M	90	PPWC	38	0.12	—	—
9.	Bickley	1974	Unpub.	K	A, P	M	18.33	PPWC	45	0.157	—	—
10.	Bird	1993	Pub.	K	A, P	M	48	PPWC	43	-0.442	—	—
11.	Blanchard	1994	Pub.	K	P	O	0.1	POWC	42	0.305	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
12.	Blanchard	1991	Pub.	K	P	O	0.1	POWC	24	0.045	—	—
13.	Bolton	1935	Pub.	K	A, P	M	—	PPWC	162	0.265	—	—
14.	Byrnes	1990	Pub.	K, E	A	O	4	PPWC	164	0.218	—	—
15.	Cadd	1994	Pub.	K	A, P	M	2	PPWC	102	0.327	—	—
16.	Cernat	2001	Pub.	K	P	O	0.1	POWC	35	0.063	—	—
17.	Curl	2002	Unpub.	K	A, P	O	1.5	PPWC	102	0.248	—	—
18.	Cushner	1989	Pub.	K	A, P	M	6	PPWC	43	0.26	—	—
19.	Czopp	2006	Pub.	K	P	O	0.1	PPWC	74	0.412	—	—
20.	Davidson	1991	Unpub.	K	A, P	M	12.5	PPWC	40	0.195	—	—
21.	Davis	1975	Pub.	K	P	O	0.33	PPWC	66	0.193	—	—
22.	Dickson	2010	Pub.	K	A, P	M	30	PPWC	60	-0.116	—	—
23.	Duck	2003	Pub.	K	P	O	0.5	PPWC	38	0.115	—	—
24.	Earnest	1968	Unpub.	K, E	A, P	M	20	PPWC	100	0.72	—	—
25.	Esses	2002	Pub.	E	P	O	0.1	POWC	40	0.41	—	—
26.	Feinman (comparison 1)	1982	Unpub.	K, S	A	M	44	PPWC	102	0.433	—	—
27.	Feinman (comparison 2)	1982	Unpub.	K, S	A	M	44	PPWC	104	0.43	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
28.	Feuchte	2010	Unpub.	K, E, S	A, P	M	36	PPWC	140	0.5	—	—
29.	Geissler	2008	Unpub.	K	A, P	M	18	PPWC	88	0.367	88	0.074
30.	Goerke	1984	Unpub.	K	A, P	M	14	PPWC	165	0.064	—	—
31.	Graves	1975	Unpub.	K	P	O	0.5	PPWC	24	-0.131	—	—
32.	Gray	1975	Pub.	K	P	O	0.83	POWC	111	0.504	87	0.244
33.	Greenberg	1957	Pub.	K	A	O	1	PPWC	82	0.154	—	—
34.	Gwinn	1998	Unpub.	K	A, P	M	11.25	PPWC	41	0.051	—	—
35.	Hamarneh	1995	Unpub.	K	A, P	O	1	PPWC	89	0.497	—	—
36.	Hawkins	1981	Unpub.	K	A, P	M	45	PPWC	182	-0.173	—	—
37.	Heath	1992	Unpub.	K	A, P	O	1	PPWC	35	0.244	—	—
38.	Hines	1984	Unpub.	K	P	M	2.72	PPWC	148	-0.099	—	—
39.	Hoffmann	2005	Unpub.	K, E	A, P	M	30	PPWC	157	0.408	—	—
40.	Hohn	1973	Pub.	E, S	A	M	9	PPWC	42	0.828	—	—
41.	Holland	1969	Unpub.	K, E	A, P	M	27.5	PPWC	60	0.588	—	—
42.	Houser	1978	Pub.	K, E	P	O	0.5	POWC	102	1.297	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
43.	Jones	1983	Unpub.	K	A, P	M	7	PPWC	258	0.01	—	—
44.	Karafantis (comparison 1)	2004	Pub.	K	P	O	1	POWC	143	0.045	—	—
45.	Karafantis (comparison 2)	2004	Pub.	K	P	O	1	POWC	152	0.143	—	—
46.	Keats	1989	Unpub.	K	A, P	M	34	PPWC	97	0.204	—	—
47.	Kehoe	1978	Pub.	E	A, P	M	4	POWC	71	0.547	—	—
48.	Kimoto	1974	Unpub.	K	A, P	M	—	PPWC	66	-0.025	—	—
49.	Klein	1963	Pub.	K, E	P	O	0.5	PPWC	814	0.602	—	—
50.	Klitzke	2005	Unpub.	K, E	A	M	20	PPWC	72	0.362	—	—
51.	Koeller	1977	Pub.	K	P	M	3	POWC	143	0.085	—	—
52.	Kraemer	1975	Unpub.	K	P	O	1	POWC	97	0.511	—	—
53.	Laghaie	1991	Unpub.	K	P	M	1	POWC	77	0.002	—	—
54.	Lemmer	2003	Unpub.	K, S	A, P	M	36	PPWC	102	0.033	—	—
55.	Levinson	1954	Pub.	K	A, P	M	45	PPWC	47	0.263	—	—
56.	Levy	2005	Pub.	K	P	O	0.5	POWC	60	0.769	—	—
57.	Litcher	1973	Pub.	K	A, P	M	11.67	POWC	100	-0.022	—	—
58.	Macphee	1994	Pub.	K	A, P	M	30	PPWC	527	0.079	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
59.	Madden	1970	Unpub.	K	A, P	M	—	PPWC	111	0.22	—	—
60.	Maluso	1992	Unpub.	K, E	A	O	0.83	PPWC	—	—	229	-0.041
61.	Maxson	1970	Unpub.	K	A, P	M	15	PPWC	107	0.478	—	—
62.	May	1997	Unpub.	K	A, P	M	11.25	PPWC	70	0.046	—	—
63.	Mcilveen	1988	Unpub.	K	A, P	M	6	PPWC	50	0.34	—	—
64.	Megumi	1994	Unpub.	K	A, P	M	3.75	PPWC	106	0.219	—	—
65.	Neto	2006	Pub.	K	A, P	M	—	PPWC	33	0.432	—	—
66.	Newell	2001	Unpub.	K, E	A, P	M	7	POWC	108	0.812	—	—
67.	Okyne	1980	Unpub.	K	A, P	M	—	PPWC	120	0.064	—	—
68.	Olson (comparison 1)	2006	Pub.	K	P	O	0.25	POWC	47	-0.4	—	—
69.	Olson (comparison 2)	2006	Pub.	K	P	O	0.25	POWC	98	-0.137	—	—
70.	Orson	1971	Unpub.	K, E	A, P	M	20	PPWC	63	1.043	—	—
71.	Page	1974	Unpub.	K, E	A, P	M	6	PPWC	26	0.895	—	—
72.	Parish (comparison 1)	1974	Pub.	K	P	O	0.5	POWC	22	0.564	—	—
73.	Parish (comparison 2)	1974	Pub.	K	P	O	0.5	POWC	28	0.76	—	—
74.	Patruno	2001	Unpub.	K, E	A, P	M	14	PPWC	55	0.747	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
75.	Peacock	1992	Unpub.	K	A, P	M	24	PPWC	44	0.072	—	—
76.	Pearl	1955	Pub.	S	A	M	60	PPWC	14	0.388	—	—
77.	Persson (comparison 1)	2003	Pub.	K	P	M	1.33	PPWC	30	-0.005	—	—
78.	Persson (comparison 2)	2003	Pub.	K	P	O	0.17	PPWC	60	-0.053	—	—
79.	Pettijohn	2009	Pub.	K	A, P	M	31.99	PPWC	59	0.755	—	—
80.	Probst	2003	Pub.	K	A, P	M	34	PPWC	94	0.15	—	—
81.	Schaefer	2004	Unpub.	K	A, P	M	18	PPWC	91	0.238	—	—
82.	Schiele	1979	Unpub.	K, E	A, P	M	14	PPWC	48	0.416	—	—
83.	Schrüfer	2003	Unpub.	K, E	A, P	O	3	PPWC	169	0.679	—	—
84.	Schwartz	1972	Unpub.	K	A, P	O	2.5	POWC	80	-0.246	—	—
85.	Shirley	1988	Unpub.	K, E	A, P	M	45	PPWC	248	0.626	—	—
86.	Shiver (comparison 1)	1970	Unpub.	K	A, P	M	27.5	PPWC	60	0.469	—	—
87.	Shiver (comparison 2)	1970	Unpub.	K	A, P	M	27.5	PPWC	60	0.215	—	—
88.	Singh	2004	Unpub.	K	P	O	0.08	POWC	110	-0.168	—	—
89.	Smey	1983	Unpub.	K, S	A, P	M	14	POWC	61	0.603	61	0.559
90.	Soble	2011	Pub.	E	P	O	0.33	PPWC	138	0.338	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
91.	Sousa	2005	Pub.	K	A, P	M	18	PPWC	193	0.373	—	—
92.	Stewart	2003	Pub.	K, E	A, P	M	20	POWC	24	0.747	—	—
93.	Tauran	1967	Unpub.	K	P	M	3	PPWC	59	-0.062	—	—
94.	Timmermans	2000	Unpub.	K, S	A, P	M	39	PPWC	112	0.124	—	—
95.	Turner (comparison 1)	2008	Pub.	K, E	A, P	M	—	PPWC	44	0.466	—	—
96.	Turner (comparison 2)	2008	Pub.	K, E	A, P	M	—	PPWC			43	0.175
97.	Vrij	2003	Pub.	K	P	O	0.08	POWC	225	0.114	—	—
98.	Wagner	2004	Unpub.	K, S	A, P	M	84	PPWC	82	0.628	56	0.232
99.	Walker	1971	Unpub.	K	P	M	9.9	PPWC	40	0.221	—	—
100.	Webster	1994	Unpub.	K, E	A, P	M	18	PPWC	21	0.619	—	—
101.	West	1997	Unpub.	K, E	A	O	8	PPWC	62	0.553	38	0.46
102.	Wilke (comparison 1)	1997	Unpub.	K	A, P	M	—	PPWC	52	0.166	—	—
103.	Wilke (comparison 2)	1997	Unpub.	K, S	A, P	M	10.41	PPWC	73	0.135	—	—
104.	Williams	1945	Unpub.	K, E	A, P	M	6	PPWC	100	0.519	100	0.06
105.	Williams	1961	Pub.	K	P	M	20	PPWC	88	0.357	—	—
106.	Yawkey	1973	Pub.	K, E	A, P	M	9.33	PPWC	104	0.415	—	—

Note. Content: K = knowledge, E = empathy, S = skills; Method: A = active, P = passive; Days: M = multi-day, O = one-day; Design: refers to the design that could be used to calculated an effect size; PPWC = pretest-posttest with control, POWC = posttest only with control; *N* = sample size; *g* = effect size.

Table B2

Characteristics of the Intervention-Control Comparisons of the Secondary Cluster

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
1.	Aoki	2000	Unpub.	K	A, P	M	22	PPSG	20	-0.064	—	—
2.	Ayres	1973	Pub.	K	A, P	M	—	PPSG	20	0.195	—	—
3.	Baba	2004	Pub.	K	A, P	M	32	PPSG	161	0.208	—	—
4.	Boatright	2005	Pub.	K, E	A, P	M	—	PPSG	114	0.097	—	—
5.	Broaddus	1986	Unpub.	K	A, P	M	9.33	PPSG	33	-0.024	—	—
6.	Brown	1996	Pub.	K, E	A, P	M	48	PPSG	35	0.082	—	—
7.	Butler	1982	Unpub.	E	A	M	45	PPSG	52	0.293	—	—
8.	Case.	2007	Pub.	K	A, P	M	30	PPSG	147	0.111	—	—
9.	Clinton	1983	Unpub.	K, E	A, P	M	28	PPSG	30	0.627	—	—
10.	Cullen	2008	Unpub.	K	A, P	M	30	PPSG	5	0.56	—	—
11.	Darigan	1991	Unpub.	K, E	A, P	M	9	PPSG	33	-0.3	—	—
12.	Dine (comparison 1)	1994	Unpub.	K, E	A, P	O	1	PPSG	23	0.904	—	—
13.	Dine (comparison 2)	1994	Unpub.	K, E	A, P	O	1	PPSG	28	-0.258	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
14.	Droba	1932	Pub.	K	A, P	M	—	PPSG	30	0.314	—	—
15.	Elley	1964	Pub.	K	A, P	M	40.5	PPSG	29	0.347	—	—
16.	Fife	1994	Unpub.	K	A, P	M	—	PPSG	21	0.093	—	—
17.	Gordon (comparison 1)	1983	Unpub.	K	P	O	0.5	PPSG	130	0.047	—	—
18.	Gordon (comparison 2)	1983	Unpub.	K	P	O	0.5	PPSG	50	0.103	—	—
19.	Gordon (comparison 3)	1983	Unpub.	K	P	O	0.5	PPSG	114	-0.016	—	—
20.	Hayes	1969	Unpub.	K, E	A, P	M	59.94	PPSG	57	0.658	—	—
21.	Hill	2001	Pub.	K, E	A, P	M	24	PPSG	62	0.519	31	0.252
22.	Hussey	2010	Pub.	K	A, P	M	30	PPSG	36	0.507	—	—
23.	Johnson	1980	Unpub.	K, S	A, P	M	8	PPSG	36	0.543	—	—
24.	Korhonen	2002	Unpub.	K	A, P	M	16	PPSG	117	0.041	—	—
25.	Kuperus	1992	Unpub.	K	A, P	M	55	PPSG	23	0.657	—	—
26.	Lefley (comparison 1)	1985	Pub.	K	A, P	M	64	PPSG	68	0.206	—	—
27.	Lefley (comparison 2)	1985	Pub.	K	A, P	M	64	PPSG	54	-0.075	—	—
28.	Levinson	1951	Pub.	K	A, P	M	45	PPSG	21	0.399	—	—
29.	Lillis (comparison 1)	2007	Pub.	K, S	A	O	1.25	PPSG	16	0.364	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
30.	Lillis (comparison 2)	2007	Pub.	K	A, P	O	1.25	PPSG	16	-0.056	—	—
31.	Lopez	1998	Pub.	K	A, P	M	30	PPSG	203	0.32	—	—
32.	Martin	2001	Unpub.	K, E	A, P	M	35	PPSG	60	0.186	—	—
33.	Mcmahon	1995	Unpub.	K	A, P	M	4.17	PPSG	19	0.15	—	—
34.	Neville	1996	Pub.	K	A, P	M	45	PPSG	29	0.094	—	—
35.	Pedersen	2009	Pub.	K	A, P	M	27	PPSG	19	0.456	—	—
36.	Pedersen	2008	Pub.	K	A, P	M	12	PPSG	62	0.739	—	—
37.	Sanchez	1997	Pub.	K, S	A, P	M	—	PPSG	79	-0.137	—	—
38.	Singh	1974	Pub.	K	A, P	M	24	PPSG	20	1.432	—	—
39.	Smith	1943	Pub.	K	A, P	M	30	PPSG	64	0.324	—	—
40.	Stangor	2001	Pub.	K	P	O	0.1	PPSG	12	0.353	—	—
41.	Steed	2010	Pub.	K	A, P	O	6	PPSG	11	0.107	—	—
42.	Stenson	1978	Unpub.	K	A, P	M	10	PPSG	61	0.016	—	—
43.	Tuttle (comparison 1)	1979	Pub.	K	P	M	1.1	PPSG	54	0.144	—	—
44.	Tuttle (comparison 2)	1979	Pub.	K	P	M	1.1	PPSG	64	0.099	—	—
45.	Van Soest	1996	Pub.	K	A, P	M	30	PPSG	92	0.148	—	—

									Direct posttest		Delayed posttest	
Nr.	First Author	Year	Publication	Content	Method	Days	Hours	Design	*N*	*g*	*N*	*g*
46.	Wieder	1954	Pub.	S	A	M	22	PPSG	27	0.335	—	—
47.	Wittig (comparison 1)	2000	Pub.	K	A	M	8	PPSG	540	-0.064	—	—
48.	Wittig (comparison 2)	2000	Pub.	K	A	M	8	PPSG	112	0.13	—	—

Note. Content: K = knowledge, E = empathy, S = skills; Method: A = active, P = passive; Days: M = multi-day, O = one-day; Design: refers to the design that could be used to calculated an effect size; PPSG = pretest-posttest single group; *N* = sample size; *g* = effect size.

Table 1

Description of the Included Intervention-Control Comparisons

		Primary cluster (k = 106)		Secondary cluster (k = 48)	
Variable	Value	k	%	k	%
Decade	Before 1961	5	4.7	4	8.3
	1961 – 1970	10	9.4	2	4.2
	1971 – 1980	23	21.7	6	12.5
	1981 – 1990	13	12.3	8	16.7
	1991 – 2000	22	20.8	14	29.2
	2001 – 2010	33	31.1	14	29.2
Type of document	Published	46	43.4	29	60.4
	Journal article	46	43.4	27	56.3
	Book / book chapter	—	—	2	4.2
	Unpublished	60	56.6	19	39.6
	Dissertation / master thesis	55	51.9	18	37.5
	Other unpublished	5	4.7	1	2.1
Country of the first author	USA	85	80.2	41	85.4
	Other	21	19.8	7	14.6
Content (Axis 1)	Knowledge	67	63.2	34	70.8
	Empathy	3	2.8	1	2.1
	Skills	1	0.9	1	2.1
	Knowledge and empathy	24	22.6	9	18.8
	Knowledge and skills	9	8.5	3	6.3
	Empathy and skills	1	0.9	—	—
	Knowledge, empathy, and skills	1	0.9	—	—

		Primary cluster (k = 106)		Secondary cluster (k = 48)	
Variable	Value	k	%	k	%
Method (Axis 2)	Passive	30	28.3	6	12.5
	Active	10	9.4	5	10.4
	Passive and active	66	62.3	37	77.1
Duration (Axis 3); one- vs. multi-day	One-day intervention	33	31.1	9	18.8
	Multi-day intervention	73	68.9	39	81.3
Duration (Axis 3); number of net	1 – 10 hours	51	48.1	18	37.5
intervention hours (categorized for	11 – 50 hours	41	38.7	21	43.8
illustration purposes)	51 – 100 hours	3	2.8	4	8.3
	Cannot be specified	11	10.4	5	10.4
Status	Majority	83	78.3	39	81.3
	Minority	7	6.6	3	6.3
	Majority and minority	3	2.8	5	10.4
	No status hierarchy	13	12.3	1	2.1
Reason for implementation	Research interests	86	81.1	34	70.8
	Practical reasons	20	18.9	14	29.2
Country of implementation	Australia	1	0.9	3	6.3
	Canada	4	3.8	2	4.2
	Finnland	—	—	1	2.1
	Germany	8	7.5	—	—
	Ghana	1	0.9	—	—
	Great Britain	4	3.8	—	—
	New Zealand	1	0.9	1	2.1
	Portugal	2	1.9	—	—
	Romania	1	0.9	—	—
	USA	84	79.2	41	85.4

		Primary cluster (k = 106)		Secondary cluster (k = 48)	
Variable	Value	k	%	k	%
Age of the participants	< 6 years	7	6.6	1	2.1
	6 – 9 years	8	7.5	4	4.3
	10 – 13 years	20	18.9	6	12.5
	14 – 18 years	21	19.8	7	14.6
	> 18 years	50	47.2	30	62.5
Sex of the participants (% female)	0 – 30%	3	2.8	2	4.2
	31 – 70%	58	54.7	22	45.8
	71 – 100%	18	17.0	11	22.9
	Cannot be specified	27	25.5	13	27.1
Design	Posttest only with control (POWC)	25	23.6	—	—
	Pretest-posttest with control (PPWC)	81	76.4	—	—
	Pretest-posttest single group (PPSG)	—	—	48	100.0
Assignment to conditions	Randomized (individuals)	40	37.7	—	—
	Not randomized	66	62.3	—	—
	No control group	—	—	48	100.0
Type of control	No treatment	79	74.5	—	—
	Placebo treatment	27	25.5	—	—
	No control group	—	—	48	100.0
Nature of posttests	Only direct (< 1 month)	97	91.5	47	98.0
	Only delayed (1 – 4 months)	3	2.8	—	—
	Direct and delayed	6	5.7	1	2.1

		Primary cluster (k = 106)		Secondary cluster (k = 48)	
Variable	Value	k	%	k	%
Interval between the end of the intervention and the direct pottest	0 day – 1 day	68	64.2	39	81.3
	> 1 day – 1 week	30	28.3	9	18.8
	> 1 week – less than 1 month	5	4.7	—	—
Interval between the end of the intervention and the delayed posttest	1 month – less than 2 months	5	55.6	—	—
	2 months – 4 months	4	44.4	1	100.0
Total sample size of the comparison (direct posttest)	Up to 30	11	10.4	20	41.7
	31 – 100	58	54.7	19	39.6
	101 – 200	29	27.4	7	14.6
	201 – 300	4	3.8	1	2.1
	> 300	2	1.9	1	2.1
Mean attrition rate (direct posttest)	Up to 10 %	71	67.0	17	34.4
	11 – 30%	13	12.3	14	29.2
	31 – 50%	3	2.8	3	6.3
	> 50 %	1	0.9	2	4.2
	Cannot be specified	18	17.0	12	25.0
Attrition rate – difference between the intervention and the control group (direct posttest)	Up to 10 %	65	61.3	—	—
	11 – 30%	4	3.8	—	—
	31 – 50%	—	—	—	—
	> 50 %	—	—	—	—
	Cannot be specified	37	34.9		
Total number of items measuring ethnic attitudes	1 – 10	13	12.3	13	27.1
	11 – 50	78	73.6	28	58.3
	51 – 100	6	5.7	2	4.2
	> 100	4	3.8	2	4.2
	Cannot be specified	5	4.7	3	6.3

		Primary cluster (k = 106)		Secondary cluster (k = 48)	
Variable	Value	k	%	k	%
DV – content	Affective/behavioral	8	7.5	14	29.2
	Cognitive	40	37.3	13	27.1
	Mixed	58	54.7	21	43.8
DV – type of measure	Likert/social distance	76	71.7	41	85.4
	Semantic differential	9	8.5	1	2.1
	Mixed	21	19.9	6	12.5
DV – level of generalization	One target outgroup	65	61.3	22	45.8
	Multiple target outgroups	8	7.5	7	14.6
	Unspecified ethnic outgroups	19	17.9	13	27.1
	Mixed (target and unspecified)	14	13.2	6	12.5

Note. DV = dependent variable.

Table 2

General Effectiveness of Information Interventions

Cluster	Time of posttest	$\hat{\mu}_\theta$	95% CI	Q	$\hat{\tau}^2$	I^2	k	N
Primary	Direct	0.28***	[0.22, 0.34]	274.26***	0.05	61.84	103	9961
	Delayed	0.16*	[0.03, 0.30]	8.43	0.01	17.47	9	782
Secondary	Direct	0.22***	[0.14, 0.29]	205.32***	0.05	80.76	48	3109

Note. $\hat{\mu}_\theta$ = estimated average of the true effects; CI = confidence interval; Q = Homogeneity statistic; $\hat{\tau}^2$ = estimated variance between the true effects; I^2 = amount of true variance among total variance; k = number of intervention-control comparisons; N = total number of participants.

* $p < .05$. *** $p < .001$.

Table 3

Bivariate Correlations Between Potential Moderators

Moderator	Active	Multi-day	Hours
Empathy	0.19	0.06	-0.03
Active		0.63**	0.51**
Multi-day			0.56**

Note. empathy, active, and multi-day are dummy coded variables, the presence of the respective aspects is indicated by 1; sample size is $k = 103$ when hours is not involved and $k = 94$ when hours is involved.

** $p < .01$.

Table 4

Results of a Weighted WLS Meta-Regression Including Hypothesized Moderators

Moderator	$\hat{\beta}$	*SE*	95% CI
Intercept	0.17**	0.05	[0.06, 0.28]
Empathy	0.37***	0.06	[0.26,0.49]
Active	0.04	0.07	[-0.10, 0.19]
Multi-day	-0.04	0.07	[-0.17, 0.10]

Note. $\hat{\beta}$ = estimate of the regression parameter; *SE* = adjusted standard error of the regression parameter; CI = confidence interval.

** $p < .01$. *** $p < .001$.

Table 5

Effectiveness of Information Interventions in the Primary Cluster (k=103) as a Function of Content, Method, and Duration

Variable	$\hat{\mu}_\theta$	95% CI	*k*	Q_{model}
Content (empathy involved)				
Empathy-evoking component	0.56***	[0.46, 0.65]	27	45.78***
No empathy-evoking component	0.18***	[0.12, 0.23]	76	
Content (configurations)				
Knowledge and empathy	0.57***	[0.47, 0.67]	22	44.55***
Knowledge and skills	0.35***	[0.18, 0.52]	8	
Knowledge	0.16***	[0.10, 0.22]	68	
Method				
Active procedure involved	0.30***	[0.24, 0.37]	73	1.21
Passive	0.23***	[0.11, 0.34]	30	
Duration				
Multi-day intervention	0.28***	[0.21, 0.35]	71	0.06
One-day intervention	0.29***	[0.18, 0.40]	32	

Note. $\hat{\mu}_\theta$ = estimated average of the true effects; CI = confidence interval; *k* = number of intervention-control comparisons; Q_{model} = test whether the average true effects differ between the levels of the moderator.

*** $p < .001$.

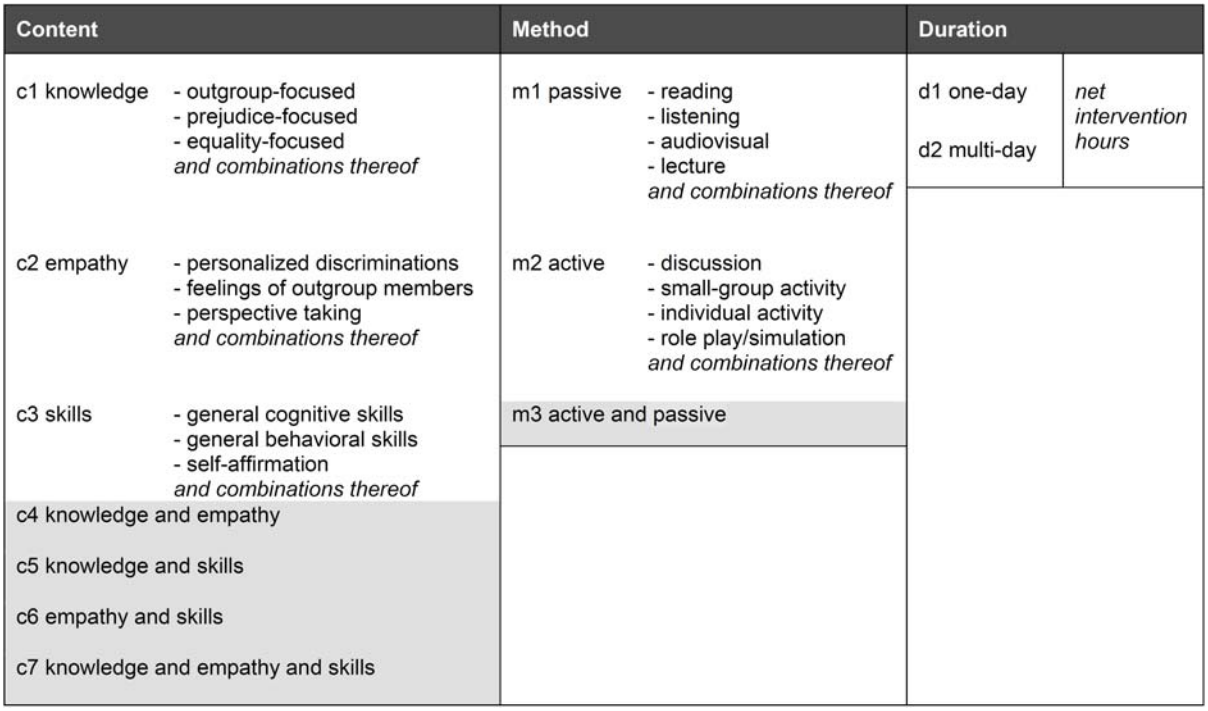

Figure 1. Illustration of a multi-axial taxonomy for the classification of information interventions. Combinations of characteristics have a grey background.

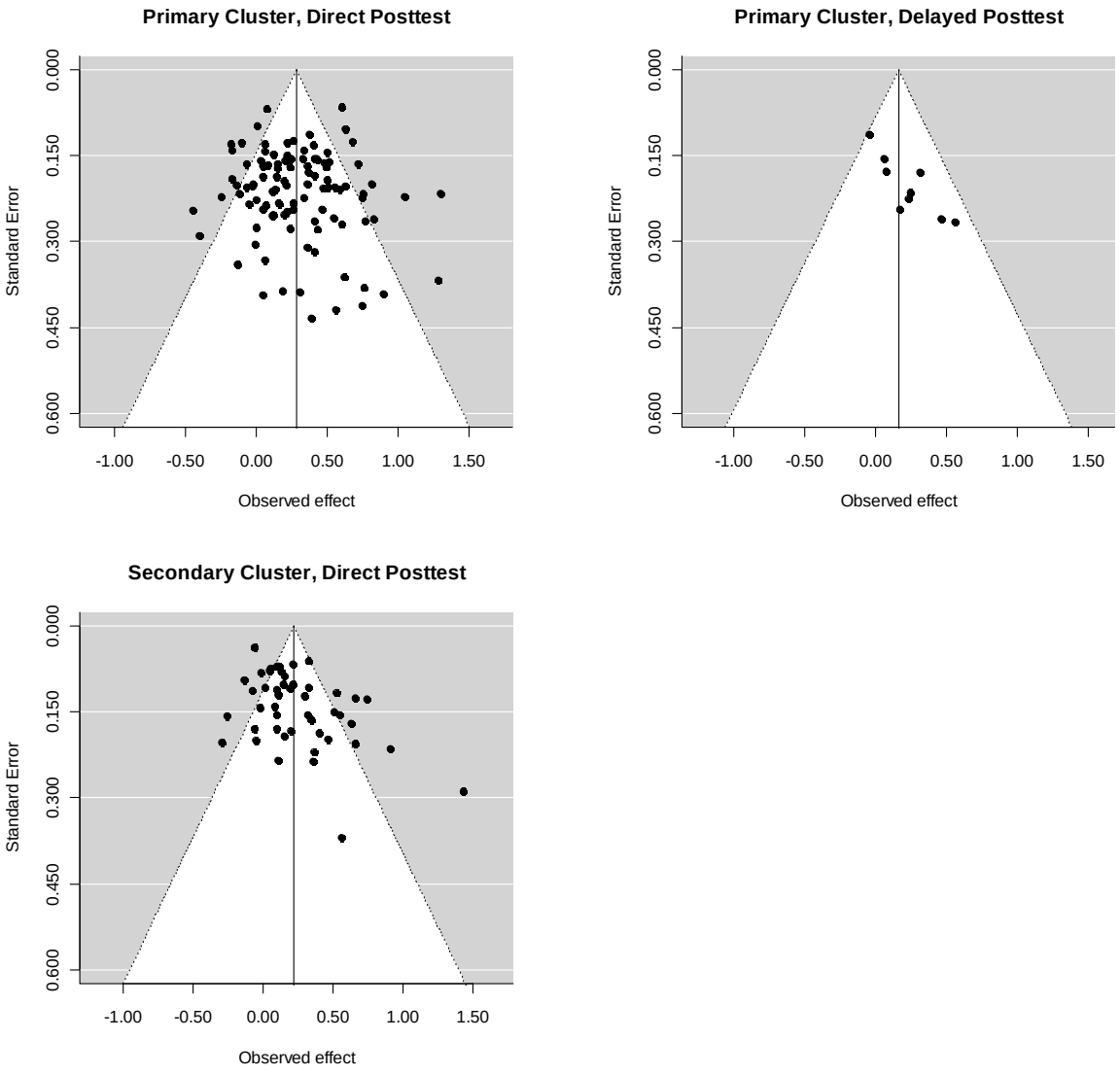

Figure 2. Funnel plot for the direct (k = 103) and delayed (k = 9) posttest in the primary cluster as well as for the direct (k = 48) posttest in the secondary cluster. The points represent the included intervention-control comparisons. On the axis of abscissae, they indicate the observed effect sizes, on the axis of ordinates, they display the corresponding standard errors. The estimated average true effect is indicated by a vertical line.

4. General Discussion

4.1. Is it Possible to Improve Ethnic Attitudes?

The manuscripts included in this dissertation describe two meta-analytic evaluations of interventions to reduce ethnic prejudice.

The first meta-analysis (Manuscript #1) was conducted in order to test the impact of interventions that are based on the intergroup contact theory and have implemented structured direct or indirect contact. The findings show that these programs typically improve ethnic attitudes with an estimated average true effect of $\hat{\mu}_\theta = 0.26$ in the primary cluster (i.e., the set of effect sizes that originated from studies with more rigorous designs) and $\hat{\mu}_\theta = 0.41$ in the secondary cluster that consists of studies that used a pretest-posttest single group design.

The research presented in Manuscript #2 addressed information interventions, i.e., programs that provide input that is not based on interethnic contact but on knowledge-, skill-, and/or empathy-based content. Before the actual meta-analysis is presented in this manuscript, we suggest a new way to classify information interventions and introduce a multi-axial taxonomy that encompasses content, method, and duration. The average estimated true effect of information programs is $\hat{\mu}_\theta = 0.28$ in the primary cluster and $\hat{\mu}_\theta = 0.22$ in the secondary cluster.

Based on these results, it can be stated that both contact and information programs typically improve ethnic attitudes and are worth implementing. In addition, the general effectiveness of these two broad approaches to reduce ethnic prejudice is comparable in size. Only the mean effect of contact studies in the secondary cluster is to some extent higher. Given the findings for the primary cluster with studies of a higher methodological quality and ignorance of all other aspects, it seems that contact and information programs are equally effective. However, such a comparison has to be made with more attention to detail (see below).

The magnitude of both effects in the primary cluster can be classified as "small to medium" (Cohen, 1988) and as "educationally relevant" (Tallmadge, 1977). At the same time, analyses of delayed posttests demonstrated that the effect of contact and information programs does not immediately diminish after the interventions have ended but lasts at least one month after the end of the initiatives.

To conclude, the impact of these two interventional approaches is not excessively large but sustained over time. Given that the aim of social interventions is to generate enduring changes, the findings obtained with my analyses are actually more encouraging than a large but rapidly decaying effect would be. Unless incidents of ethnic discrimination that otherwise would occur in the time period immediately following the intervention could be prevented, the latter would be of no special use.

Besides the general effectiveness of contact as well as of information programs, we have respectively postulated detailed hypotheses and tested them. With the first meta-analysis, we demonstrate that contact programs are more effective for groups that have a higher status position (e.g., European Americans) in the context of the respective contact program than for ethnic minority groups (e.g., African Americans). However, interaction-based initiatives have a positive outcome for ethnic minorities, too. In addition, we show that contact interventions are also effective in regions that are marked by a (former) intractable conflict (e.g., in the Middle East, Cyprus), that not only direct (i.e., face-to-face) but also indirect (i.e., non-face-to-face) variants of intergroup contact reduce ethnic prejudice, and, finally, that the effect of contact strategies is not restricted to persons who had been met in the specific situation but also generalizes to the outgroup as a whole as well as to unspecified ethnic outgroups (i.e., when the utilized instruments measure attitudes toward ethnic outgroups in general without focusing on specific groups).

In reference to information interventions, we demonstrate with the second meta-analysis that programs including empathy-evoking content are more effective than those that do not address empathy. Furthermore,—and in regard to future implementations probably even more important—we unveiled a sequence of the impact of different information initiatives. Strategies that only deal with the enhancement of knowledge (e.g., by explaining historical-cultural backgrounds of ethnic outgroups) have a positive outcome. However, interventions that additionally focus on the sophistication of social-cognitive skills have a higher impact, and, in turn, programs that address both knowledge and empathy are the most effective. According to this finding, we clearly recommend for future information interventions to not only focus on knowledge but to add further content. In this context, it also has to be mentioned that some content-configurations (i.e., only empathy; only skills; empathy and skills; knowledge, empathy, and skills) could not be included in this analysis because of small case numbers. Little is known about the impact of these configurations. Additionally, with regard to the content axis, we found that the beneficial effect of including empathy-evoking content also consists when solely cognitive operationalizations of ethnic prejudice were used. Moreover and independent from the content axis, we expected that the impact of information programs is positively affected by the inclusion of at least one active procedure (i.e., discussion, small group work to generate a product, individual activity, and role play/simulation game) to pass over the respective content and by prolonged treatment (i.e., by using a multi-day instead of a one-day implementation and, respectively, by using a strategy that has a larger number of net intervention hours). Though, the findings indicate that the two aspects do not moderate the outcome of information interventions. We conducted further analyses to be able to reveal influences of these variables that may be concealed when considered isolatedly. Even with these analyses, no support for a moderating function of utilized methods as well as of program length was found. Therefore, it can be concluded in

regard to information interventions that content matters—these programs are particularly effective when empathy is addressed—and that the type of the used methods as well as the length of the initiatives has no substantial influence on their outcome. Furthermore, no interactions between the examined domains exist, meaning that the positive contribution of empathy-focused content does not depend on the methods that were generally applied by the programs and on the duration of the program. Please note that this conclusion is not identical with the more specific conclusion that the method that was used to address empathy does not affect the impact of empathy. Unfortunately, the latter aspect could not be investigated within the context of the work presented in Manuscript #2 since an exact assignment and relational coding in the sense of "which method(s) were used to pass over which content" was often not possible.

To summarize, ethnic prejudice can be reduced by means of contact programs as well as by the use of information interventions. In this context, one final word: The findings presented in this dissertation are of great importance since the two meta-analyses are based on extensive literature searches, cumbersome retrievals of fulltexts that are not easy to access, and on precise assessments of several thousands of documents in order to extract those intervention-control comparisons that were finally included. These efforts are reflected in the considerable portion of included unpublished research. This fact is very beneficial as it is typically assumed (see Rothstein, Sutton, & Borenstein, 2005) that the process of publication systematically favors studies with significant positive findings. Therefore, an exclusive or primarily consideration of published work would substantially threaten the validity of the estimations of the average true effects. Interestingly and with regard to both manuscripts, comparisons that were conducted between published and unpublished work did not reveal significant differences. This might signal that the typically assumed problem of publication bias is not of major concern in the field of social interventions that is considered here.

4.2. A Comparison of Contact and Information Interventions

Having characterized contact and information programs as well as summarized the impact of these general categories of strategies to improve ethnic relations, two central issues arise: To what extent can these approaches be conceptually differentiated and the results we obtained with the two meta-analyses be compared. I will elaborate on these issues starting with the first, that is, a conceptual comparison between contact and information interventions. Subsequently, I will deal with the second issue, that is, the extent to which a comparison of their mean effects is possible.

In reference to the conceptualizations we use in Manuscript #1 and #2, one might ask whether contact and information programs can be accurately separated from each other. This question would be more straightforward to answer some years ago before the concept of indirect intergroup contact had been introduced.

Although some contact interventions—to be more precise those contact programs we label contact meetings—may also contain (a limited amount of) informational input to reduce ethnic prejudice, most of them do not, at least not systematically. The sub-type of contact programs that is termed cooperative learning methods typically includes informational input, however, the content of it usually does not address interethnic relations but rather the academic curriculum (e.g., mathematics). In addition, whereas it can not be ruled out that occasional interethnic contact takes place within the context of information programs, it is not introduced by purpose. This clearly distinguishes them from the category of contact programs. According to our definition given in Manuscript #1, the latter have to implement contact in a highly structured way in order to assure that not only opportunities for contact are established but that interethnic interactions really take place. Therefore, direct contact programs and information-based interventions can be clearly separated.

However,—as already mentioned—with the introduction of variants of indirect contact further elaboration is required in order to be able to decide whether or not contact and information programs can be conceptually separated. As described in Manuscript #1, a parasocial contact hypothesis (Schiappa, Gregg, & Hewes, 2005; see also Davis, 2008; Schiappa, Gregg, & Hewes, 2006) was stated in the field of communication science. It is characterized by the authors as a "communication analogue to Allport's (1954) contact hypothesis" (Schiappa et al, 2005, p. 92) and they claim that unidirectional "contact" between consumers and mass-mediated information from outgroup members will improve attitudes toward the respective outgroup. To accept parasocial "contact" as a further version of (indirect) contact would make it nearly impossible to conceptually differentiate between contact programs and other interventions to improve ethnic attitudes. In my opinion, a consideration of the process that is specified with the parasocial contact hypothesis (Schiappa, Gregg, & Hewes, 2005; see also Davis, 2008; Schiappa, Gregg, & Hewes, 2006) as contact is not warranted. Parasocial "contact" lacks one essential element that all other forms of direct and indirect contact include: bi-directional interactions between members of different groups. Direct contact programs are based on face-to-face intergroup interactions, extended contact refers to the knowledge of intergroup friendships or positive intergroup interactions, imagined contact addresses the mental simulation of interactions with an outgroup member, and, finally, the variant of indirect contact I label virtual contact refers to bi-directional contact that is mediated by computer technology. In contrast, parasocial "contact" is an one-directional, information-based communication from an outgroup member to a receiver without containing any possibility of bi-directionality. However, this does not imply that mass mediated content cannot qualify as extended contact. When it explicitly and purposely displays instances of interethnic friendships or—as a more liberal conceptualization—positive interethnic interactions, it is extended contact. To conclude, indirect contact and

informational approaches can be clearly differentiated by stating that informational input does not qualify as indirect contact as long as it does not explicitly depict positive interactions between members of different groups.

Having demonstrated that contact and information interventions can be conceptually distinguished, it is now of interest to shed light on whether a direct comparison of the magnitude of the average true effects we identified for the two approaches is meaningful.

As presented, the estimated average true effect of contact interventions in the "better" primary cluster is $\hat{\mu}_\theta = 0.26$, the corresponding impact of information programs is estimated to be $\hat{\mu}_\theta = 0.28$. These estimates are of approximately equal size, but can it be concluded that contact and information initiatives are generally equally effective, or put in another way: is this comparison fair? In regard to this, two aspects are worth mentioning. The first one refers to conceptual aspects and the second one to the empirical results.

First, contact interventions may be in a disadvantaged position as they probably have to be seen in most cases as providing an increment of contact that is added to the contact that already exists and could have already improved ethnic attitudes. Consequently, the effect size of contact initiatives may represent the amount of additional improvement that is added in the context of the thin air that remains after intergroup attitudes have been already improved via the same route by pre-interventional contact in daily life activities. Therefore, programs based on the principle of interethnic interactions may have a higher effect when implemented without or with only negligible prior contact. Although information strategies are typically realized under a similar contact-baseline condition, it can be assumed that they have an advantage as they utilize other ways of changing intergroup attitudes (e.g., explicitly addressing empathy with informational input). To sum up this line of theoretical reasoning, while the two types of interventions have similar effect sizes, in a setting where there is no or only a very low amount of pre-interventional contact, contact programs might be more

effective than information strategies. Maybe this issue can be systematically addressed by future work.

Secondly, as already described, contact programs have a higher impact for ethnic majorities than for minorities. Majority samples have an estimated average true effect of $\hat{\mu}_\theta$ = 0.37 which exceeds the general estimated mean effect of information interventions ($\hat{\mu}_\theta$ = 0.28). The impact of information initiatives does not change when only those 83 comparisons are included that consist of at least 75% majority members. Accordingly, it can be concluded that when comparisons are considered that—in the case of contact studies—exclusively or—in the case of information studies—primarily contain majority members, contact interventions may be more effective. In the total sample of the meta-analysis of information interventions, seven comparisons include at least 75% minority members. The estimated average effect for this sub-group is $\hat{\mu}_\theta$ = 0.39 and exceeds the estimated mean effect of $\hat{\mu}_\theta$ = 0.16 for minority-based comparisons in the meta-analysis of contact studies. Although the estimated average true effects are not directly comparable since the contact-based comparisons consist exclusively of minority members and the information-based comparisons of at least 75% minority members, they indicate a clear trend. The presented data suggest that—when ignoring the sub-types of interventions—for ethnic majorities contact programs may be more effective than information interventions and for minority persons information strategies may be more beneficial than contact initiatives.

4.3. Future Directions

Meta-analyses are a very powerful research instrument, they summarize—when conducted in an exhaustive manner—the existing research on a given question by the use of quantitative methods. Therefore, they can provide general and aggregated information about main effects (as expressed in the estimated average true effect) and the existence or non-

existence of moderators of the magnitude of the effect sizes. Besides this main aim of meta-analyses, a useful by-product of the comprehensive and systematized processes of literature search, assessment of the literature, and coding of the studies is that they deliver a very fruitful overview of the current state of research on the topic. By doing so, they can identify research gaps and, in turn, stimulate new primary (and secondary) research which than can be integrated in an updated or new meta-analysis that has consequently—with a larger sample of studies or, probably even more important, with larger numbers of cases in crucial moderator-related cells—more options to analyze the data. In the light of this process of mutual enhancement that I here suggest, the two meta-analyses have—as mentioned in Manuscript #1 and Manuscript #2—detected some research gaps. I will describe the corresponding aspects hereinafter. Following this section, I will also address things the primary (and secondary) research can do for meta-analyses to make them less resource consuming.

4.3.1. Suggestions for Future Primary (and Secondary) Research

In the context of the two meta-analyses, I detected several insufficiencies of the current state of research. Referring to this, the following issues are worth mentioning.

1. According to the integrated threat theory (Stephan & Stephan, 2000), prejudice toward an outgroup is elicted by factors such as intergroup anxiety (i.e., concerns in intergroup situations about being embarrassed, rejected, or ridiculed), realistic threats (i.e., threats to the existence of the ingroup, e.g., regarding political or economic power), and symbolic threats (i.e., threats to the worldview and moral values of the ingroup. Existing evidence supports this approach (e.g., Riek, Mania, & Gaertner, 2006; Stephan, Renfro, Esses, Stephan, & Martin, 2005; Tausch, Hewstone, & Roy (2009).

Although the reduction of intergroup anxiety and threat can be seen as a probable by-product of the interventions we found and integrated, none of them dealed explicitly with it.

Therefore, since this approach is promising, interventions have to be designed and evaluated that explicitly address the influencing factors suggested by Stephan and Stephan (2000).

2. Some other interventions have been already designed but only rarely evaluated or only evaluated with special groups. In this context, as mentioned in Manuscript #1, variants of indirect contact interventions for ethnic minorities have to be systematically evaluated. Maybe these programs open up an opportunity to improve the effectiveness of contact strategies for minority samples. Concerning this, a sequence of initial indirect contact and following direct contact might be an especially beneficial approach.

In regard to information interventions, programs have to be implemented and evaluated that explicitly address all of the three content components we differentiate: knowledge, empathy, and skills. We hardly found such studies. Our results show that initiatives that address knowledge and empathy are more effective than programs that deal with knowledge and skills which, in turn, have a better outcome than strategies that only address knowledge. Consequently, it can be expected that a combination of all three aspects is particularly effective.

3. In Manuscript #2, We have tested whether the duration of an information intervention does influence its effectiveness. We did not find such an effect. Although not explicitly mentioned in Manuscript #1, our meta-analysis of contact programs did also not reveal a significant influence of program length. In this context, we discuss in Manuscript #2 the possibility that shorter interventions may be more intensive and therefore might compensate a possible disadvantage that is due to their limited length. Since—as also described—it is not possible in a meta-analysis to determine the actual intensity of social interventions with the information given in the documents, the assumed interplay between duration and intensity has to be investigated by future primary or secondary research. This

aspect is of great importance because it could help to distribute given budgets in a way that is the most beneficial.

4. In both meta-analyses, we found that the available evaluations in almost all cases do not systematically test whether the programs really work via the path that is suggested or whether the outcome could be (partly) also the result of another, unexpected mechanism that is triggered by the intervention. Therefore, I (see also Lipsey & Cordray, 2000) recommend to not only focus on the program at the one side and on its impact at the other, but to also test mediating processes. As long as no systematization is given concerning the specific processes in the field of actual interventions, categorizations of programs with our multi-axial taxonomy or other classification systems have to focus on the content that is described in the documents, disregarding the circumstance whether a given program that contains knowledge-enhancing, empathy-focused and/or skill-associated content really improves knowledge, empathy, and/or skills and by reason of it, in turn, reduces prejudice or whether other factors are involved.

5. With the two meta-analyses that are described in Manuscript #1 and #2, we originally intended to not only consider indices of explicit ethnic attitudes but also of actual, not self-reported behavior toward ethnic outgroups as well as of implicit ethnic attitudes. However, the evaluations that could be included did not focus on these variables that are of great interest. Future research has to consider dependent variables that clearly address these aspects. When such studies exist, they can be systematically integrated in an updated meta-analysis that can test whether or not contact and information programs do really reduce actual ethnic discrimination.

6. Finally, we were able to test whether the effect of contact and information rapidly fades after the intervention has ended or whether it is sustained over a certain period of time. The available data support the latter. However, with regard to the two primary clusters that

encompass research of higher quality, only a few studies evaluated a prejudice reduction intervention with a delay of one or more months. In addition, in the meta-analysis of information programs, there was no study with a posttest that acquired data more than four months after the end of the intervention. In the contact meta-analysis, some studies used a longer delay between the end of treatment and an additional posttest, but no evaluation included a posttest that was conducted more than one year after the program. Therefore, further primary research investigating long-term effects is needed. With a larger sample of studies and longer delays, more systematic conclusions regarding the persistence of the improved ethnic attitudes can be drawn.

4.3.2. What the Primary (and Secondary) Research can do for Meta-Analyses

I have already mentioned that primary and secondary research on the one hand and meta-analyses on the other should stimulate and support each other. During the process of conducting the meta-analyses that are described in Manuscript #1 and #2, I was confronted with several time and resource consuming obstacles that can be avoided by routinely providing the following information in research reports (that describe evaluations of social interventions).

1. Adequate description of the independent variable (i.e., the intervention). Most studies provide basic information regarding the program that was implemented. Therefore, I could code the characteristics that are specified in my multi-axial taxonomy. However, some documents do not contain a sufficient description, so that the person who is conducting a meta-analysis has to enter the time consuming route of searching for additional information. I suggest that the following intervention characteristics should be routinely described in a given research report with respectively at least some sentences and—if possible—a concretization at the operational level: initiation (why was the program designed and why

was it implemented in the specific case?); theoretical background (who designed the program and with which theoretical assumptions and backgrounds?); content (what was done?); method (how was it done?), duration (how long was it done?); a clear statement regarding the existence or non-existence of implementation problems and special incidents. This list is far from being complete, but in light of the fact that space is often limited, it represents an adequate trade-off.

2. Adequate description of the dependent variable(s). In case a study applies an instrument that is well known in an unmodified fashion, no detailed information is needed. In contrast, when often used instruments are modified or self-constructed instruments are utilized, detailed information concerning the content and the psychometric quality (validity and reliability) of the concrete operatinalization is needed. At best—if possible—, a complete list of the used items is presented so that the meta-analysist can accurately assess its content and decide whether the respective instrument is in line with his/her inclusion criteria or not. Regarding this matter, just displaying one or very few sample items can be misleading.

3. Provision of group-based sample characteristics. Research reports should routinely contain a group-based description of the included participants, at least concerning the basic socio-demographic variables age, sex, and ethnic background. Although most documents provide data concerning these variables for the total sample, many do not include basic socio-demographic statistics separated by treatment and control group(s). These more fain-grained data are, however, important to be able to code study characteristics in a more specific fashion, to test for moderating influences with these more accurate codings, and, of course, to assess the comparability between the intervention and the control group.

4. Provision of descriptive statistics in a group- and time-based manner. The issue touched here is by far the most problematic and annoying one for a person who conducts a meta-analysis. Sometimes, a study satisfies each inclusion criteria but does not allow

calculating an effect size without entering the field of speculation (see Manuscript #1 and #2 for details). This very often initiates a process of trying to obtain additional data, for instance by personal conversation with the author, which, in turn, is often not successful in delivering the required data, especially when the study was conducted some time ago. In order to open up the person who is conducting a meta-analysis all possibilities to calculate effect sizes, the following descriptive statistics should be provided, at best in a tabular manner: means, standard deviations, and sample sizes; respectively separated by treatment and control groups as well as by time of measurement (when repeated measurements were realized, pretest-postest correlations should be additionally provided—if available).

Although this list resulted from the meta-analyses that were conducted in the context of this dissertation, I assume that the aspects I describe are not restricted to it. A routine inclusion of the corresponding data in research reports would help to conduct meta-analyses faster and with less required resources.

.

References

Cohen. J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Davis, Y. A. (2008). *The parasocial contact hypothesis: Implications for changing racial attitudes* (Unpublished master's thesis). Ohio State University.

Lipsey, M. W., & Cordray, D. S. (2000). Evaluation methods for social intervention. *Annual Review of Psychology, 51*, 345–375.

Riek, B. M., Mania, E. W., & Gaertner, S. L. (2006). Intergroup threat and outgroup attitudes: A meta-analytic review. *Personality and Social Psychology Review, 10***,** 336–353.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: Wiley.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2005). The parasocial contact hypothesis. *Communication Monographs, 72*, 92–115.

Schiappa, E., Gregg, P. B., & Hewes, D. E. (2006). Can one TV show make a difference? Will & Grace and the parasocial contact hypothesis. *Journal of Homosexuality*, *51*, 15–37.

Stephan, W. G., Renfro, C. L., Esses, V. M., Stephan, C. W., & Martin, T. (2005). The effects of feeling threatened on attitudes toward immigrants. *International Journal of Intercultural Relations, 29*, 1-19.

Stephan, W. G., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 23–45). Mahwah, NJ: Erlbaum.

Tallmadge, G. K. (1977). *The joint dissemination review panel idea book*. Washington, DC: National Institute of Education, and the US Office of Education.

Tausch, N., Hewstone, M., Kenworthy, J. B., Psaltis, C., Schmid, K., Popan, J. R., Cairns, E., et al. (2010). Secondary transfer effects of intergroup contact: Alternative accounts and underlying processes. *Journal of Personality and Social Psychology, 99*, 282–30

5. Zusammenfassung

Meta-Analytische Evaluationen der Effektivität von Interventionen zur Verbesserung ethnischer Einstellungen

Alltägliche Mediendarstellungen, eigene Beobachtungen und Forschungsbefunde (siehe zum Beispiel Klink & Wagner, 1999) zeigen, dass Menschen nur aufgrund ihres ethnischen Hintergrundes benachteiligt werden (können). Vorurteile, d.h. negative Einstellungen (zum Beispiel Antipathie) gegenüber Menschen, die anderen ethnischen Gruppen angehören werden dabei üblicherweise als ein Fundament solcher Diskriminierungen angesehen.

Sozialpsychologische Theorien können erklären, warum es zu negativen gruppenbasierten Einstellungen kommt. In diesem Rahmen wurden mehrere Einflussfaktoren wie soziale Kategorisierungs- und Identitätsprozesse (siehe Tajfel & Turner, 1979, 1986) sowie das Empfinden von Bedrohung (siehe Stephan & Stephan, 2000) identifiziert. Ein bekannter sozialpsychologischer Ansatz geht dabei das Problem gruppenbasierter Vorurteile von der anderen Seite an. Nach der Kontakttheorie (siehe Allport, 1954; Pettigrew & Tropp, 2011) werden Intergruppenbeziehungen verbessert, wenn Mitglieder unterschiedlicher Gruppen miteinander interagieren. Dieser Ansatz wird durch Fragebogenuntersuchungen gut belegt. Erweiterungen der ursprünglichen Kontaktannahme beinhalten, dass Intergruppenbeziehungen beispielsweise auch durch die Darstellung interethnischer Freundschaften (z.B. mit Filmen oder Bildergeschichten) verbessert werden können, wenn eine Person selbst also gar nicht am Kontakt teilnimmt, sondern ihn nur beobachtet.

Sozialpsychologische Ansätze wie die Kontakttheorie lassen sich nutzen, um Interventionen abzuleiten, die im Praxisfeld (beispielsweise in der Schule) eingesetzt werden können um ethnische Vorurteile abzubauen. Neben kontaktbasierten Maßnahmen werden im Praxisfeld oft auch solche Programme realisiert, die ethnische Vorurteile beispielsweise durch die Vermittlung neuen Wissens (zum Beispiel über die Kultur anderer ethnischer Gruppen),

durch die Förderung von Empathie (beispielsweise durch eine gezielte Aufforderung, die Perspektive einer Person einzunehmen, die einer ethnischen Minderheit angehört) oder durch die Förderung allgemeiner Kompetenzen (zum Beispiel im Sinne eines Training des logischen Schlussfolgerns) reduzieren möchten. Diese zweite große Gruppe von Maßnahmen wird in Abgrenzung zu Kontaktprogrammen hier als Informationsmaßnahmen bezeichnet.

Es stellt sich die Frage, ob die beschriebenen Programme ihr Ziel, also die Verbesserung ethnischer Einstellungen, auch wirklich erreichen. Daher ist ihre Wirksamkeit mit Ergebnisevaluation zu überprüfen. Dabei wird in der Regel das gemittelte Vorurteilsniveau einer Teilnehmergruppe mit dem einer Kontrollbedingung verglichen. Hat die Trainingsgruppe einen günstigeren Wert, also weniger Vorurteile, kann unter bestimmten Voraussetzungen darauf geschlossen werden, dass die Maßnahme, wie intendiert, auch wirklich Vorurteile abbaut. Ergebnisevaluationen helfen also zu beurteilen, ob sich die Investition in die Durchführung eines Programms überhaupt lohnt.

In der Literatur sind mittlerweile sehr viele derartiger Evaluationen dokumentiert. Deshalb stellt sich von einer übergeordneten Perspektive aus betrachtet die Frage, wie wirksam solche Maßnahmen generell (also nicht nur im Einzelfall) sind und welche Programme die besten Effekte haben.

Ein methodisches Verfahren zur Untersuchung solcher Fragestellungen ist die Meta-Analyse. Es handelt sich dabei um einen Forschungsprozess, in dem möglichst alle Studien zur jeweiligen Forschungsfrage gesammelt, einzeln ausgewertet und ihre Ergebnisse schließlich numerisch zusammengeführt werden. Das Verfahren der Meta-Analyse aggregiert dabei die Befunde der einzelnen Studien und liefert einen Gesamtwert der Effektivität. Darüber hinaus lässt sich auch untersuchen, welche Aspekte der Einzelstudien (zum Beispiel Art der Intervention oder Dauer des Programms) einen Einfluß auf den Gesamtwert haben.

Ziel meiner Arbeit war, das Verfahren der Meta-Analyse anzuwenden, um die Wirksamkeit von Maßnahmen zur Reduktion ethnischer Vorurteile zu untersuchen. Dabei habe ich zwei getrennte Meta-Analysen durchgeführt.

Die erste Meta-Analyse hat solche Programme systematisch untersucht, die auf der Kontakttheorie basieren. Ich habe insgesamt 121 Kontraste analysiert, die jeweils eine Kontaktbedingung mit einer Kontrollbedingung verglichen haben. Der Gesamtwert zeigt, dass Kontaktprogramme ethnische Vorurteile reduzieren und, dass der Effekt zeitlich stabil ist. Zudem habe ich spezifische, weiterführende Fragestellungen untersucht. Dabei wurde gefunden, dass Kontaktprogramme für Mitglieder der ethnischen Mehrheit effektiver sind als für Mitglieder ethnischer Minderheiten. Zudem verdeutlichen die Ergebnisse, dass Kontaktprogramme selbst in solchen Gebieten wirksam sind, die durch einen ernsthaften Konflikt zwischen verschiedenen Ethnien gekennzeichnet sind (beispielsweise im Nahen Osten). Außerdem zeigte sich, dass nicht nur direkte (also „face-to-face“-basierte) sondern auch indirekte Kontaktmaßnahmen effektiv sind. Die Befunde sprechen zudem eindeutig dafür, dass nicht nur die Einstellungen gegenüber denjenigen Personen verbessert werden, mit denen der Kontakt stattgefunden hat, sondern die verbesserten Einstellungen auf die gesamte Gruppe generalisiert werden.

In einer zweiten Meta-Analyse habe ich Informationsprogramme untersucht. Also Interventionen, die Wissen vermitteln wollen, Empathie fördern möchten oder beabsichtigen, allgemeine sozial-kognitive Fähigkeiten zu trainieren. Im Rahmen dieser Analyse wurden insgesamt 154 Vergleiche untersucht. Es zeigt sich über alle Studien hinweg ein positiver Gesamteffekt der hinsichtlich seiner Höhe vergleichbar mit der Gesamtwirksamkeit von Kontaktprogrammen ist. Darüber hinaus habe ich untersucht, ob bestimmte Merkmale zu einer größeren Effektivität führen. Die Ergebnisse sprechen dafür, dass insbesondere solche Programme effektiv sind, die empathiefördernde (siehe oben) Inhalte besitzen. Entgegen meiner Erwartungen hing der Gesamterfolg der Maßnahmen nicht davon ab, ob im jeweiligen

Programm auch aktive Methoden (wie beispielsweise Kleingruppenarbeiten oder Rollenspiele) enthalten waren. Auch wurde die Wirksamkeit der Interventionen überraschenderweise nicht von der Dauer der Maßnahmen beeinflusst.

Die Ergebnisse beider Meta-Analysen zeigen, dass ethnische Einstellungen gezielt verbessert werden können. Dabei sind sowohl Kontaktprogramme als auch Informationsinterventionen wirksam. Demgegenüber hängt der Erfolg einer Maßnahme zur Vorurteilsreduktion nicht von ihrer Länge ab (dies gilt auch für Kontaktprogramme). Es scheint auch so zu sein, dass der gezielte Einsatz von aktiven, partizipatorischen Techniken keinen bedeutsamen Einfluß auf die Programmwirksamkeit hat. Diese Aspekte sollten jedoch mit weiterführenden Untersuchungen genauer überprüft werden.

Literatur

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

Klink, A., & Wagner, U. (1999). Discrimination against ethnic minorities in Germany: Going back to the field. *Journal of Applied Social Psychology, 29*, 402–423.

Pettigrew, T. F., & Tropp, L .R. (2011). *When groups meet: The dynamics of intergroup contact*. London, UK: Routledge/Psychology Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–48). Monterey, CA: Brooks/Cole.

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

Stephan, W. S., & Stephan, C. W. (2000). An integrated threat theory of prejudice. In S. Oskamp (Ed.). *Reducing prejudice and discrimination* (pp. 23–45). Mahwah: Erlbaum.

6. Danksagung

Ich danke allen Personen sehr herzlich, die mir bei der Fertigstellung dieser Arbeit geholfen haben und mich untersützt haben.

Ich danke meinem Erstanleiter, Ulrich Wagner, für sein Engagement sowie für die wertvollen und richtungsweisenden Gespräche. Sehr wichtig war, dass er auch für spontane Nachfragen immer zur Verfügung stand.

Herzlich bedanken möchte ich mich auch bei Mario Gollwitzer für seinen Einsatz sowie für wichtige Hinweise und Anregungen. Auch er stand für Nachfragen immer zur Verfügung.

Mein Dank gilt ebenfalls Peter Schmidt, der mir in mehrerern Gesprächen wichtige Hinweise gegeben hat.

Darüber hinaus möchte ich mich bei allen Mitgliedern der Arbeitsgruppen Methodenlehre und Sozialpsychologie sowie des Graduiertenkollegs Gruppenbezogene Menschenfeindlichkeit für Anregungen, Unterstützung und sehr nette Gespräche (auch abseits der Arbeit) bedanken. Zudem möchte ich mich bei allen studentischen Hilfskräften bedanken, die mir bei der Jagd nach Studien, deren Durchsicht sowie bei Korrekturen der Artikel behilflich waren.

Schließlich gilt mein Dank meinen Eltern Erich und Annelen, meinem Bruder Torben und Sabine Völker. Auch bedanken möchte ich mich bei Andreas Rühl.

7. Angaben zur Person

Gliederungspunkt 7 enthält persönliche Daten. Er ist deshalb nicht Bestandteil der Online-Veröffentlichung.

8. Erklärung

Ich versichere, dass ich meine Dissertation

“Meta-Analytic Evaluations of Interventions to Improve Ethnic Attitudes”

selbstständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

_____ _____

(Ort/Datum) Gunnar Lemmer