Preventing prejudice and improving intergroup attitudes: A meta-analysis of child and adolescent training programs

Andreas Beelmann*, Kim Sarah Heinemann

Department of Research Synthesis, Intervention, and Evaluation, Institute of Psychology, Friedrich-Schiller-University of Jena, Germany

A R T I C L E   I N F O

Article history:
Received 29 August 2012
Received in revised form 1 November 2013
Accepted 7 November 2013
Available online 4 January 2014

Keywords:
Prejudice
Intergroup attitudes
Meta-analysis
Intervention
Children
Adolescents

A B S T R A C T

This article reports the results of a meta-analysis of 81 research reports containing 122 intervention-control comparisons of structured programs to reduce prejudice or promote positive intergroup attitudes in children and adolescents. Overall, the analysis revealed a mean effect size of $d = 0.30$, indicating low to moderate intervention effects. From the great variety of different approaches, interventions based on direct contact experiences among social-cognitive training programs designed to promote empathy and perspective taking showed the strongest effect sizes. In addition, effects varied according to the program participant’s social status (higher effects for majority groups), the target out-group (lower effect sizes for ethnic vs. disabled and aged out-groups), and the type of outcome assessment (higher effects for cognitive vs. affective and behavioral measures of intergroup attitudes). The discussion considers several limitations including the lack of implementation and follow-up research as well as future direction of research on promoting intergroup relations.

© 2013 Elsevier Inc. All rights reserved.

Prejudice and other forms of negative intergroup attitudes cause serious social problems in many societies throughout the world. Even in their mildest form, the consequences for human beings may include social exclusion and segregation, health and behavioral problems, poorer chances on the labor market, and even a generally more negative quality of life (see, e.g., Paradies, 2006; Williams, Neighbors, & Jackson, 2003). Previous social-developmental research has indicated clearly that prejudice and other forms of biased intergroup attitudes are not just restricted to adulthood, but that children start to favor their own social group as soon as the basic processes of social categorization and identification emerge in early childhood (see Levy & Kilien, 2008; Raabe & Beelmann, 2011).

Although definitions of prejudice have changed over the last decades (see Duckitt, 2010), it can generally be viewed as “any attitude, emotion, or behavior toward members of a group, which directly or indirectly implies some negativity or antipathy toward that group” (Brown, 2010, p. 7). According to this definition, prejudice may have different manifestations and is multifaceted in nature. On a global level, one can first differentiate between in-group preference (e.g., ascribing more positive characteristics to our own social group) and out-group derogation (ascribing more negative characteristics to the social group we do not belong to). These seem to be distinct constructs, although they have common negative consequences for out-group members (Brewer, 1999). In addition, negative attitudes also have a multifaceted character and are normally divided into a cognitive component (attributing negative characteristics such as being mean or aggressive to members of the out-group), an emotional component (e.g., liking them less), and a behavioral component (e.g., exhibiting negative behavior such as social exclusion). In line with the broad range of different operationalizations, recent social-developmental research has studied prejudice within a broader framework of intergroup attitude development. This has led to the identification of important individual and social developmental factors and processes. For example, it has been proposed that cognitive and social-cognitive abilities such as classification skills and perspective taking (Aboud, 1988; Bigler & Liben, 2007), the development of social identity (Nesdale, 2004), or moral decision making (Kilien & Rutland, 2011) are crucial contributors to intergroup attitude development in children. Other research has focused more on social variables such as intergroup contact and cross-group friendships (Davies, Tropp, Aron, Pettigrew, & Wright, 2011), social norms (Rutland, 2004), and intergroup threat (Bar-Tal & Teichman, 2005; Riek, Mania, & Gaertner, 2006).

However, although all this research documents major scientific progress in understanding the emergence of prejudiced attitudes, there is an ongoing debate over which are the most important individual and social factors that need to be addressed in systematic interventions designed to prevent negative intergroup attitudes and associated problems of intolerance and discrimination in intergroup relations (Kilien, Rutland, & Ruck, 2011; Tropp & Mallett, 2011). As a result, rather diverse intervention programs have been developed and tested during the last 30 years (e.g., Oskamp, 2000; Stephan & Stephan, 2001). The present meta-analysis summarizes research on
the effectiveness of standardized psychological and educational intervention programs to prevent or reduce prejudice or otherwise promote positive intergroup attitudes and relations in childhood and adolescence. It studies the general effectiveness as well as specific effects indicating which program will be the best alternative for whom and on what outcome measure.

Interventions aiming to promote intergroup attitudes and relations have a long history and can be traced back to the integrative school systems movement in the United States (Brown v. Board of Education, 1954) and to Allport’s (1954) classic work on the psychological nature of prejudice. Both initiatives generated a great deal of empirical research (see Pettigrew & Tropp, 2006; Schofield, 1995; Schofield & Hausmann, 2004), particularly in the social context of schools and universities, and this has led to a variety of intervention concepts. Other approaches place more emphasis on education in political values (e.g., democracy), multicultural diversity, morality, and values such as equality, respect, and tolerance (Stephan & Vogt, 2004). Despite their different histories and disciplinary roots, all these approaches are based on the assumption that prejudice and negative intergroup attitudes are associated with a number of social problems and phenomena within our societies such as the social exclusion and discrimination of ethnic minorities, immigrants, handicapped, or other groups; racism, hate crimes, and right-wing extremism; or even international conflicts. Therefore, preventing prejudice and negative intergroup attitudes is assumed to be an appropriate tool for avoiding or at least reducing these problems and generally promoting more positive intergroup relations, tolerance, and justice within our societies.

In an earlier review of this research with children and adolescents, Aboud and Levy (2000) distinguished between five groups of interventions ranging from integrated schooling, bilingual education, multicultural education, promoting social-cognitive skills, to training in role-playing and empathy. They classified these as being based on three different theoretical concepts about intergroup attitudes (see also Killen et al., 2011). The first most prominent theoretical concept is founded on the intergroup contact hypothesis (Allport, 1954; Pettigrew & Tropp, 2011). This states that contact between members of different social groups leads to less prejudiced attitudes and generally promotes positive intergroup relations. This effect of intergroup contact is even more pronounced when the situation complies with the criteria of positive contact, that is, an equal status of members of both social groups within the situation, support for the contact from authorities, and cooperation in achieving a joint goal. The intergroup contact hypothesis is one of the best supported theories in social psychology with hundreds of studies showing that the effects apparently hold for a variety of different situations, settings, and samples (see Paluck & Green, 2009; Pettigrew & Tropp, 2006, for reviews). Well-known examples of this intervention type are integrated schooling, cooperative learning techniques, or even public campaigns with high-profile individuals from the social out-group. In addition, recent extensions of intergroup contact theory have led to programs based on the idea of extended, that is, indirect contact—when someone gets to know an in-group member who is in contact with an out-group member (see Wright, Aron, McLaughlin-Volpe, & Ropp, 1997).

The second theoretical construct refers to general socialization theory and social learning theory. This assumes that social experiences, knowledge, and information about the social world and social out-groups as well as knowledge about relevant concepts in the context of intergroup attitudes (e.g., information about prejudice, democratic values, and cultural diversity) lead to more positive intergroup relations. Multicultural or antibias training programs are good examples for these knowledge- and information-based interventions.

Finally, the third theoretical construct is social-cognitive developmental theory. In general, it assumes that children’s intergroup attitudes often reflect their stage in sociocognitive development and that the increasing development of distinct sociocognitive abilities leads to a decreasing tendency to have biased attitudes. This forms the basis for trainings in cognitive and social-cognitive skills (e.g., classification skills, social categorization, perspective taking, conflict resolution, moral decision making) that interrelate with the development of intergroup attitudes.

Aboud and Levy’s (2000) classification of interventions into three different theoretical concepts is, of course, neither independent nor exhaustive. Because they are grounded in diverse scientific disciplines such as social development, psychology, multicultural education, sociology of migration and diversity, criminal justice, or politics, there are several ways to classify interventions in this field. For example, Paluck and Green (2009) recently presented an extended review of prejudice reduction interventions that distinguished between categories such as cooperative learning, entertainment strategies using books or films, discussion and peer influence, and instruction. Other authors such as Killen et al. (2011) have differentiated between interventions to promote intergroup attitudes by influencing peer relations, adult-child interaction, and social-cognitive judgments. In a recently published article, Aboud et al. (2012) differentiated between contact-based, media/instruction-based, and antibias/multicultural interventions. These and other examples illustrate that interventions in this field could possibly be best characterized as a mixture of programs with diverse underlying theories, goals, contents, strategic concepts, and intervention methods.

Perhaps as a consequence of this diversity in classifying intergroup interventions, the outcomes of systematic evaluations in this field have revealed no clear empirical consensus on what should be viewed as the best way to prevent prejudice and promote positive intergroup attitudes among children and adolescents (Aboud & Levy, 2000; Paluck & Green, 2006). However, several meta-analyses and systematic reviews point to some promising approaches. For example, Pettigrew and Tropp’s (2006) comprehensive meta-analysis on the intergroup contact hypothesis showed that programs based on contact between members of different social groups seem to be of central value in influencing intergroup attitudes independent of age, type of attitudes, cultures, and social conditions. Aboud et al. (2012) confirmed these results even for young children under the age of 8. Likewise, several educational techniques seem to offer promising ways of reducing prejudice. For example, intensive studies on the effects of cooperative learning techniques (Johnson & Johnson, 1989, 2000; Roseth, Johnson, & Johnson, 2008) have shown that they markedly increase not only school and academic achievement but also interpersonal attraction and relations between members of different ethnic groups in the classroom compared with individual and competitively oriented learning strategies.

Other approaches such as multicultural, diversity, and antiracism training programs have a somewhat more limited empirical foundation, but also seem to produce low to moderate, but nonetheless positive, effect sizes on prejudice—at least in adult populations (Paluck & Green, 2009; Stephan, Renfro, & Stephan, 2004; Verkuyten, 2010). Empirical evaluations of cognitive and social-cognitive programs are also basically positive, showing that a reduction in biased judgment processes and prejudice can be achieved (Aboud & Levy, 2000; Paluck & Green, 2009). Nonetheless, especially in this field, any final conclusion is difficult to reach, because this label covers a wide variety of different approaches such as programs to improve cognitive skills (e.g., Bigler & Liben, 1992), social perspective taking and empathy (e.g., bystander interventions, see Stephan & Finlay, 1998), moral development (e.g., value self-confrontation, see Grube, Mayton, & Ball-Rokeach, 1994), and conflict resolution (Johnson & Johnson, 1998).

In sum, a variety of measures have been applied to improve intergroup relations and prevent prejudice. However, most of the aforementioned and other reviews (see Beelmann, 2009) do not focus specifically on psychological and educational programs with children and adolescents (Paluck & Green, 2009; Pettigrew & Tropp, 2006), are limited to ages 8 and below and do not use a meta-analytic approach (Aboud et al., 2012), are narrative in nature, or somewhat outdated (Aboud &
Therefore, the present meta-analysis aims to integrate existing research on the effectiveness of structured intervention programs designed to promote positive intergroup attitudes and prevent prejudice in children and adolescents—indeed, independent of the underlying theoretical concept (e.g., contact- or knowledge/information-based, promoting individual social-cognitive competencies).

Alongside the question of the general effectiveness of the selected programs, there are also several differential research questions. First, because different intervention concepts exist, we are naturally interested in the specific effects of different types of programs such as multicultural trainings or training in moral reasoning, different intervention components such as information about the out-group versus teaching empathy, and the influence of implementation parameters such as the number of sessions or the type of administrator. Second, because Raabe and Beelmann (2011) found a characteristic developmental course of ethnic prejudice along with significant moderators of this course, we shall analyze several possible moderators of effectiveness including age, social status (majority vs. minority), and target out-group (ethnicity, disabilities, etc.). Finally, because prejudice and intergroup attitudes are clearly multifaceted constructs (Rutland, Killen, & Abrams, 2010) with divergent operationalization and assessment strategies, we shall also analyze the effectiveness of the programs in terms of different measures of intergroup attitudes and prejudice.

Method

Study selection

Primary studies were selected according to the following eligibility criteria: First, the study had to evaluate a standardized psychological or educational intervention program aimed at reducing prejudice or otherwise improving intergroup attitudes and relations on the basis of intergroup contact, information/knowledge acquisition, or promotion of individual social-cognitive competencies. The main feature for selecting programs was their standardization, that is, there had to be at least a recognizable stepwise order of different units. Such units do not necessarily mean different program sessions, but rather different components or methodologies used in a structured manner. For this reason, we did not include interventions based solely on contact between members of an in-group and out-group with no further components (such as structured discussions about contact experience). Likewise, we excluded research on broadband educational approaches (such as integrated schooling) or special learning methods (e.g., cooperative learning techniques), because they are not constructed as standardized and structured intervention programs. In addition, contact-based interventions and cooperative learning techniques have already been investigated and summarized sufficiently well in recent years (see Pettigrew & Tropp, 2006; Roseth et al., 2008). Second, the study had to contain an intervention group and at least one untrained control group. Third, the age of the youngsters had to be below 18 years. Fourth, the outcome assessment had to contain some kind of intergroup attitudes (e.g., prejudice, in-group favoritism, attitudes toward discriminating behavior, or tolerance). Fifth, results had to be reported in a manner that permits a reliable calculation of effect sizes. Finally, sixth, research reports had to be written in English, German, French, Spanish, Italian, or another major European language.

Literature search

We applied several strategies to identify relevant evaluation studies. First, we examined the reference lists of existing reviews and meta-analyses on the modification of intergroup attitudes and on intervention programs to reduce or prevent prejudice in children and adolescents (Aboud & Levy, 2000; Graves, 1999; Levy, 1999; Levy & Killen, 2008; McGregor, 1993; Paluck & Green, 2009; Pettigrew & Tropp, 2006; Pfeifer et al., 2007; Stephan & Finlay, 1999; Stephan et al., 2004). Second, we intensively scanned electronic databases such as PsycINFO, ERIC, PSYNDEX, and ProQuest Dissertations and Theses with the combination of three groups of keywords with and without truncations (*): (a) intervention/prevention/evaluation/effect*/training, (b) child*/adolescent*, and (c) intergroup attitude*/prejudic*/ingroup bias/discriminat*/tolera*. Third, we analyzed the references given in previously identified primary studies for further relevant publications (cross-referencing). Finally, we conducted an Internet search to look for additional promising evaluation studies.

In the first round, these strategies yielded a total of 214 research reports that we considered to be relevant. We excluded 71 reports (33.1%) because they did not meet our first eligibility criterion on the program character. We dropped further 31 reports (15.4%) because a control group was missing (second eligibility criterion); and 10 reports (4.7%), because the sample was older than 18 years of age (third eligibility criterion). Finally, we excluded 15 reports (7.0%) because of missing data for the effect-size calculation and another 6 reports (2.8%) that could not be retrieved. Thus, we ended up with 81 research reports that met our eligibility criteria. Of these 81 reports, only slightly more than one-half (56%) had already been integrated in the above-mentioned reviews and meta-analyses with the highest coverage rates being 56% (Paluck & Green, 2009) and 24% (Pettigrew & Tropp, 2006).

Some of the research reports contained more than one study and/or more than one intervention or control group, therefore permitting multiple intervention–control group comparisons. As a result, our final database (and unit of analysis) for this meta-analysis was 122 intervention–control group comparisons.

Coding and computation of effect sizes

The coding scheme contained characteristics of the publications (e.g., author, year of publication), intervention programs (e.g., target out-group, content component, program parameter, duration), sample and participants (e.g., age, gender, social status), methods (e.g., design, sample size), outcome assessment (e.g., content and type of dependent measures), and, finally, study results (see Tables 1–4 for a detailed list). The second author and a trained advanced student used this scheme to code these reports and comparisons. They met periodically to discuss and clarify critical coding decisions. In addition, 30 intervention–control group comparisons (24.6%) were randomly selected and coded by an independent rater to calculate the interrater reliability of the coding sheet.

Table 1

Description of the 81 research reports.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Coding</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1958–1970</td>
<td>4</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>1971–1980</td>
<td>23</td>
<td>28.4</td>
<td></td>
</tr>
<tr>
<td>1981–1990</td>
<td>18</td>
<td>22.2</td>
<td></td>
</tr>
<tr>
<td>2001–2010</td>
<td>16</td>
<td>19.8</td>
<td></td>
</tr>
<tr>
<td>Journal article</td>
<td>51</td>
<td>63.0</td>
<td></td>
</tr>
<tr>
<td>Dissertation</td>
<td>26</td>
<td>32.0</td>
<td></td>
</tr>
<tr>
<td>Book chapter</td>
<td>2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Conference paper</td>
<td>2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Publication language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>78</td>
<td>96.3</td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>2</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Country of intervention</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>60</td>
<td>74.1</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>6</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>Great Britain</td>
<td>4</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td>3</td>
<td>3.7</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>1</td>
<td>1.2</td>
<td></td>
</tr>
</tbody>
</table>

Note. n = number of research reports.
Overall, this interrater reliability was satisfactory (M = 93.2%), ranging from 78.3% (program parameter) to 100% (e.g., country of conduct, mean age of the sample).

We used Cohen’s (1988) $d$ to compute unified effect sizes. When relevant data were available, we computed the effect sizes as the difference between the pre- and posttest scores in the intervention and
the control group divided by the pooled standard deviation of the inter- 
vention and control group at the pretest. When only posttest scores 
were available, we used these to compute effect sizes. When no means 
and standard deviations had been reported, we used recomputation 
can results with no 

tests 9 93 7.20 .150 .13 a 33 27.0 0.24 0.20 a 

Knowledge 9 3.4 0.48 0.54 abcd 9 7.7 0.44 0.46 

Social distance 36 13.5 0.21 0.22 c 24 19.7 0.20 0.20 b 

In-group favoritism 47 17.7 0.01 0.03 cabc 25 20.5 0.09 0.10 acc 

Behavioral 66 24.8 0.23 0.19a 43 35.2 0.27 0.29a 

Not specifiedb 7 2.6 0.21 0.28c 

Visual 60 22.6 0.24 0.22b 28 23.0 0.41 0.35 

Type of instrument 

Questionnaires 149 56.0 0.36 0.32a 82 67.2 0.38 0.35a 

Tests 98 37.2 0.15 0.13c 33 27.0 0.24 0.20a 

Interview 16 6.0 0.16 0.16c 8 6.6 0.17 0.10c 

Observation 2 0.8 0.00 0.00c 2 1.6 0.00 0.00c 

Stimulus material for category activation 

Lexical 151 56.8 0.24 0.22a 82 67.2 0.26 0.25 

Visual 60 22.6 0.24 0.22b 28 23.0 0.41 0.35 

Lexical and visual 20 7.5 0.35 0.29 12 9.8 0.41 0.33 

Not specifiedc 35 12.8 0.38 0.34ab 

Note. n ES = number of effect sizes. k = number of comparisons. d = unweighted effect size. d w = weighted effect size based on the random effects model. Weighted mean effect sizes within the same characteristic category sharing a common subscript differed significantly from each other. A positive effect size indicates a successful intervention (i.e., reduced prejudice or improved intergroup attitudes). 

a Explanations on coding categories are explained in Table 3. 

b Not specified measures were not integrated on comparison level. 

c Effect size did not differ significantly from zero (p > .10).
kinds of social-cognitive skill promotion). In addition, we distinguished the programs according to several program parameters (e.g., methods such as group discussion and role play, materials used, use of intervention manuals, or implementation within a regular curriculum). Table 2 shows that 93 of the 122 interventions (76.2%) were based theoretically on socialization/knowledge acquisition (by using mostly information on some social out-group), 66 (54.1%) on intergroup contact theory (indirect and direct contact), and 45 (36.9%) on social-cognitive developmental theory (mostly with perspective taking/empathy training as content). A total of 74 interventions (60.7%) used a combination of these approaches—the most frequent being a combination of socialization/knowledge acquisition and intergroup contact (k = 42, 36.1%) followed by a combination of socialization/knowledge acquisition and social-cognitive development (k = 18, 14.8%). The combination of intergroup contact and social-cognitive development was used in six interventions (4.9%). Eight studies tested the results of interventions combining all three elements (6.6%). Of the implementation components, the most frequently used methods were group discussion (k = 86, 70.5%) and printed materials (k = 91, 74.6%).

The program duration ranged from a single meeting to 44 weeks of intervention with the majority lasting 8 weeks or less (k = 71, 58.2%). The number of sessions also showed a broad range from one single session (k = 9, 7.4%) to 44 sessions (k = 2, 1.6%), although most programs contained 10 sessions or less (k = 71, 58.2%). Combined with the duration of each session, the majority of programs were of only low to medium intensity (k = 69 or 56% of programs lasted up to 10 h) with a relatively high number of studies with an unspecified dosage. The mean age of the participants varied from 3 years 6 months to 12 years (M = 10.2 years, SD = 3.67). However, there was a clear focus on elementary school children aged 7 to 10 years (see Table 2). A mixed sample of boys and girls was used in 44.2% of the comparisons (k = 54), and in 69.7% of the comparisons, the participants were majority children and adolescents (k = 85), whereas only four comparisons addressed minority groups (k = 33 comparisons included mixed samples or delivered no information on the social status of the sample).

The study design quality ranged from post-control-only studies to randomized experiments (see Table 2), with the highest frequency for pre–post-control studies with demonstrated comparability between experimental and control group (k = 70, 57.4%). Sample sizes ranged from 16 to 1,480 subjects with 64.8% of the samples containing 100 subjects or less (k = 79).

**Overall intervention effects**

Post-intervention comparison effect sizes (k = 122) ranged between d = −0.78 and d = 2.14 with 8.2% of the comparisons having negative effect sizes (k = 10). The unweighted overall mean comparison effect size was d = 0.33 (k = 122). When the random-effects model (FEM) was applied, the overall effect became slightly smaller (d = 0.27). Because the FEM revealed significant heterogeneity, Q(df = 121) = 345.39, p < .001, the random-effects model (REM) seemed to be more appropriate for our data, especially in view of the diversity of the integrated interventions. According to the REM, the overall effect was d = 0.30 (k = 122). However, the mean overall effect differed significantly from zero (p < .001) for both the FEM and the REM.

**Publication bias**

A potential threat to the interpretation and validity of any meta-analytic result is the problem of publication bias (Beelmann & Lipsey, in press; Rothstein, 2008). Publication bias refers to the tendency for only studies reporting a significant result to actually get published. The reasons for such a bias may be manifold (e.g., publication guidelines of journals, researchers’ opinions about “good,” i.e., significant data), but the effect is the same: an over-representation of significant results in published studies and therefore an overestimation of effect sizes.

To address the issue of potential publication bias, we used the funnel plot technique to display the distribution of unweighted contrast effect sizes by sample size (Light & Pillemer, 1984). Fig. 1 shows the distribution of the effect sizes by their standard errors that serve as an indicator for the precision of the effect-size estimation. Because the funnel plot was not fully symmetrical, as indicated by the 95% confidence interval around the mean weighted effect size, we conducted a regression analysis using the sample size as the independent variable and the unweighted effect sizes as the dependent variable (Macaskill, Walter, & Irwig, 2001). This allowed an informed testing of the asymmetry of the funnel plot. The regression slope (β = −10, p = .25) did not differ significantly from zero, indicating that there was no significant prediction from sample size to effect size. This cast doubt on the assumption that the observed mean effect size in this meta-analysis was distorted due to publication bias—possibly because of the relatively high number of unpublished reports (e.g., dissertations, conference paper) in this dataset (see Table 1). Indeed, a comparison between results based on published versus unpublished reports revealed at least a statistical trend in favor of published results, d = 0.34, k = 82 for published, and d = 0.23, k = 40 for unpublished reports; Q(df = 1) = 2.66, p = .10, using the REM.

**Effects on different intergroup attitude measures**

The studies used a great variety of outcome measures to assess intervention effects. Table 3 reports the definitions and most important codings of the outcome measures according to the intergroup attitudes dimension and the type of intergroup attitude measure. Table 4 presents an overview of the effect sizes based on these and additional codings of the outcome assessment on the dependent variable level (nES = 266) and comparison level (k = 122) on which effect sizes of the same category were averaged within comparisons before being integrated between comparisons. Studies focused mostly on the cognitive dimension of intergroup attitudes (nES = 162, k = 107; 60.9 and 87.7%, respectively), and prejudice measures were the main type of outcome (nES = 140, k = 94; 52.6 and 77.0%, respectively). Effects were assessed mainly by questionnaires (nES = 149, k = 82, 50.6 and 67.2%, respectively), and the stimulus material addressing the social out-group was activated most often lexically (nES = 151, k = 82; 56.8 and 67.2%, respectively).

Because categories were, of course, not independent on the comparison level (i.e., the same comparison was usually represented in more than one category), we ran the moderator analyses at the dependent variable level. For all moderator analyses, we selected the REM of effect-size integration because of the aforementioned heterogeneity of the overall effect size. These analyses revealed that the intergroup...
attitudes dimension was a significant moderator, \( Q(df = 3) = 11.41, p = 0.009 \), with outcome measures focusing on the cognitive dimension of intergroup attitudes yielding significantly higher effect sizes than outcome measures focusing on the behavioral and affective dimensions. In particular, outcomes on the affective dimension of intergroup attitudes showed effect sizes that did not differ significantly from zero on both analytic levels. Also, the type of intergroup attitude was a significant moderator, \( Q(df = 5) = 33.25, p < 0.001 \), with knowledge about the out-group and prejudice measures yielding the highest, and measures for in-group bias the lowest and nonsignificant effect sizes (see Table 4). Finally, the type of instrument used for the outcome measure accounted significantly for effect-size variability, \( Q(df = 4) = 19.55, p < 0.001 \), with questionnaires yielding significantly higher effect sizes than test scores. The stimulus material used for the activation of the out-group category was not a significant moderator, \( Q(df = 3) = 3.70, p = 0.30 \).

Moderator analyses on program and sample characteristics

Before analyzing the influence of program and sample characteristics on effectiveness, we first looked at the influence of the study design quality and the sample size to control for the influence of methodological variables on effect-size variability. Although it seemed as if poor study design (only post-control design) and low sample size (below 50) correlated negatively with effect sizes (see Table 2), neither the study design quality nor the sample size were significant moderators, \( Q(df = 3) = 0.78, p = 0.86 \) and \( Q(df = 3) = 2.80, p = 0.42 \), respectively.

Turning to program characteristics, the analysis of the target out-group revealed at least a tendency for the target out-group to be a potential moderator, \( Q(df = 2) = 5.33, p = 0.07 \). Programs addressing intergroup attitudes toward persons with disabilities yielded the highest effect size and were even significantly more effective than programs addressing intergroup attitudes toward members of other ethnicities (see Table 2).

Effect sizes for the three different theoretical orientations (socialization/ knowledge acquisition, intergroup contact, social-cognitive development) were almost identical, with \( d \) ranging from 0.31 to 0.33. However, results differed for the various content components, especially in the area of intergroup contact and social-cognitive development interventions (see Table 2). The highest effect sizes were found for interventions with direct contact elements \( (d = 0.43, k = 22) \) and training in perspective taking/empathy \( (d = 0.44, k = 24) \), whereas the lowest and nonsignificant effect sizes emerged for programs with training components in classification/social categorization \( (d = 0.14, k = 18) \) and social problem-solving skills \( (d = 0.18, k = 11) \). Of course, as already mentioned, the majority of programs used a combination of contents (see above). Therefore, we compared interventions according to their complexity and found a nearly significant difference between unimodal (e.g., only one content aspect; \( d = 0.24, k = 48 \)) and multimodal programs, \( (e.g., \) more than one content component, \( d = 0.35, k = 74) \); \( Q(df = 1) = 3.41, p = 0.06 \). However, there were no differences between several types of content combinations (e.g., knowledge acquisition/intergroup contact vs. knowledge acquisition/social-cognitive development).

In addition to these analyses, we tested whether interventions differed in their effects on various outcome measures and compared the programs according to different intergroup attitudes dimensions. In general, there were only very few differential effects. The most notable were, first, the relatively high effect of programs with a social-cognitive developmental foundation on cognitive measures of intergroup attitudes \( (d = 0.47, k = 40) \). Second, was the above-average effect sizes of programs with a direct contact component and programs covering training in perspective taking/empathy on both cognitive measures \( (d = 0.45, k = 17 \) and \( d = 0.64, k = 21, \) respectively) and behavioral measures of intergroup attitudes \( (d = 0.38, k = 5 \) and \( d = 0.45, k = 13, \) respectively). Finally, this pattern could be found generally in multimodal programs that had systematically higher effect sizes on each intergroup attitude dimension, especially on behavioral measures, on which they differed significantly from unimodal programs \( (d = 0.29, k = 25 \) vs. \( d = 0.12, k = 18) \); \( Q(df = 1) = 4.63, p = 0.03 \).

Effect sizes for the different program parameters varied between \( d = 0.23 \) and \( d = 0.35 \). Interventions using role taking/role play as a method yielded the highest effect sizes, but effect sizes in general did not differ very much from each other (see Table 2). However, because the program components and program parameters were usually con founded, we conducted a meta-regression controlling for the mutual influence of these program variables (see Table 5). Overall, this regression model was significant, \( Q(df = 16) = 38.90, p < 0.001 \), leaving no proportions of variance unexplained, \( Q(df = 105) = 110.33, p = 0.34 \). In detail, the \( \beta \) coefficients for the program components direct intergroup contact and perspective taking and the program parameters active trainer and implementation within the curriculum revealed (at least as a tendency) that these variables were significant positive moderators; in other words, the inclusion of these components and parameters yielded higher effect sizes. In contrast, programs using group discussion and printed materials yielded lower effect sizes than programs not including these components (see Table 4).

No further characteristics of the program \( (duration, number of sessions, intensity rating, trainer) or the subjects \( (age group, gender, in- group) \) accounted significantly for effect-size variability (all \( p > 0.15 \), see Table 2). Thus, none of these characteristics were significant moderators, although some single categories differed significantly from each other \( (e.g., \) programs implemented by research assistants/supervised students had higher effect sizes than programs implemented by study authors; see Table 2). In addition, effect sizes for some single categories did not differ significantly from zero (i.e., if the target in-group was minority children, \( d = 0.17 \)), indicating that there was no significant intervention effect for these groups.

Description of research reports including follow-up data

In addition to posttest data, 10 of the 81 research reports included follow-up results. These reports contained 13 intervention-control group comparisons of which 5 addressed intergroup attitudes toward persons of other ethnicities, 5 toward persons with disabilities, and 3 toward the aged (see Table 6 for further details). The follow-up assessment

<table>
<thead>
<tr>
<th>Program components</th>
<th>( \beta )</th>
<th>( p )</th>
<th>( B )</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socialization/knowledge acquisition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information on groups</td>
<td>−0.15</td>
<td>.135</td>
<td>−0.11</td>
<td>[−0.25, 0.01]</td>
</tr>
<tr>
<td>Information on prejudice</td>
<td>−0.09</td>
<td>.432</td>
<td>−0.07</td>
<td>[−0.26, 0.11]</td>
</tr>
<tr>
<td>Information on values and norms</td>
<td>.07</td>
<td>.604</td>
<td>0.07</td>
<td>[−0.21, 0.35]</td>
</tr>
<tr>
<td>Intergroup contact</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect contact</td>
<td>.17</td>
<td>.105</td>
<td>0.12</td>
<td>[−0.03, 0.27]</td>
</tr>
<tr>
<td>Direct contact</td>
<td>.25</td>
<td>.012</td>
<td>0.24</td>
<td>[0.05, 0.43]</td>
</tr>
<tr>
<td>Social-cognitive skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classification/social categorization</td>
<td>−.22</td>
<td>.029</td>
<td>−0.21</td>
<td>[−0.40, −0.02]</td>
</tr>
<tr>
<td>Perspective taking/empathy</td>
<td>.36</td>
<td>.001</td>
<td>0.33</td>
<td>[0.14, 0.53]</td>
</tr>
<tr>
<td>Social problem solving</td>
<td>−.18</td>
<td>.016</td>
<td>−0.22</td>
<td>[−0.47, 0.04]</td>
</tr>
<tr>
<td>Moral development</td>
<td>−.01</td>
<td>.953</td>
<td>−0.01</td>
<td>[−0.34, 0.32]</td>
</tr>
<tr>
<td>Program parameter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group discussion</td>
<td>−.20</td>
<td>.040</td>
<td>−0.15</td>
<td>[−0.29, −0.01]</td>
</tr>
<tr>
<td>Role-taking/role playing</td>
<td>−.03</td>
<td>.795</td>
<td>−0.02</td>
<td>[−0.18, 0.14]</td>
</tr>
<tr>
<td>Printed materials</td>
<td>−.32</td>
<td>.003</td>
<td>−0.26</td>
<td>[−0.42, −0.09]</td>
</tr>
<tr>
<td>TV-material</td>
<td>−.04</td>
<td>.644</td>
<td>−0.03</td>
<td>[−0.17, 0.11]</td>
</tr>
<tr>
<td>Manual used</td>
<td>−.02</td>
<td>.862</td>
<td>−0.02</td>
<td>[−0.20, 0.17]</td>
</tr>
<tr>
<td>Active trainer</td>
<td>.20</td>
<td>.055</td>
<td>0.20</td>
<td>[−0.00, 0.40]</td>
</tr>
<tr>
<td>Curriculum implementation</td>
<td>.17</td>
<td>.108</td>
<td>0.12</td>
<td>[−0.03, 0.26]</td>
</tr>
</tbody>
</table>

Note. \( B \) = unstandardized regression coefficient. CI = confidence interval. Positive \( \beta \) = inclusion of component increases mean effect size, exclusion decreases it. Negative \( \beta \) = inclusion of component decreases the mean effect size, exclusion increases it.
varied from 2 weeks to 16 months with a mean of about 4 months. The follow-up effect sizes ($k = 13$) ranged between $d = -1.11$ and $d = 2.30$ with two negative effect sizes (see Table 6). The unweighted overall mean comparison effect size was $d = 0.32$. Following the FEM, the overall follow-up effect decreased slightly, $d = 0.29$, $p < .001$, and again, showed significant heterogeneity, $Q(12) = 74.44$, $p < .001$. Therefore, the random-effects model seemed to be more appropriate for our follow-up data, too. According to this model, the overall follow-up effect was $d = 0.29$, $p = .05$.

Discussion

The present meta-analysis integrated the effects of standardized intervention programs designed to prevent and reduce prejudice or otherwise improve intergroup attitudes in children and adolescents. The interventions proved to be generally effective with a mean effect size of around $d = 0.30$. These outcomes are largely comparable to those effects ascertained within other prevention fields such as antisocial behavior or the outcomes for social and emotional training programs in general (see, e.g., Beelmann & Raabe, 2009; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). According to the binominal effect-size display, this could be considered to be an improvement in intergroup attitudes of around 15% (Lipsey & Wilson, 2001).

Although this effect size is viewed typically as a small to medium effect (Cohen, 1988), one should recognize that most of the programs under review are of low intensity and cost, and even small effects could have impressive implications and practical significance—for example, if slightly improved intergroup attitudes lead to less discrimination or prevent violent behavior toward out-group members (see Ellis, 2010).

However, we have to consider at least two restrictions to our results. First, the mean effect sizes must be evaluated against the background that a large proportion of the studies summarized in the present meta-analysis used proximal outcome assessments. Such assessments usually have strong ties to the content of programs and generally result in larger effect sizes (Beelmann, 2011). Therefore, the present results reflect more of the upper limit of program effectiveness. Second, the majority of primary studies assessed only immediate effects on intergroup attitudes, and only relatively few studies evaluated long-term developments or the stability of outcomes. Nonetheless, these few long-term evaluations confirmed that durable positive effects are not only possible but sometimes even higher than short-term changes (e.g., Beelmann, Saur, Ziegler, Diener, & Noack, 2010). The problem of insufficient investigations into the long-term effects of programs is well known in intervention research in general. It could perhaps be viewed as the most pressing issue when it comes to justifying prevention measures, because, in most cases, long-lasting effects (e.g., on tolerance attitudes, hate crimes, or right-wing extremism) are the intention (Beelmann, 2011). However, apart from this open question regarding effect stability and long-term effects, the present results can be viewed as reliable estimators of effectiveness, because neither publication bias nor significant confounds with methodological variables seem to restrict the validity of the data.

Therefore, the results of this meta-analysis imply first of all, that prejudice prevention and the promotion of intergroup attitudes are possible via psychological and educational training programs in children and adolescents. Nonetheless, the moderate effect sizes suggest that this approach is in need of further development and should be viewed as one way of dealing with problems of prejudice and negative intergroup relations (Aboud et al., 2012; Killen et al., 2011). Second, as in many intervention fields, program outcomes are moderated by several variables involving the assessment method, program components, and target groups. This makes it necessary to take a closer look at what works, for whom, and under which conditions.

Effects on different outcome dimensions

Many meta-analyses have shown that characteristics of the outcome measures are important effect-size moderators (Lipsey & Wilson, 2001). This could be endorsed here as well, especially with regard to the intergroup attitude dimension. Apparently, the cognitive and behavioral dimensions of intergroup attitudes are easier to influence than the emotional dimension (e.g., sympathy). This result holds for all types of intervention, and it contradicts some previous findings reporting that emotional indicators tend to show stronger effects than cognitive indicators—at least for contact-based interventions (Tropp & Pettigrew, 2009).

However, our findings should be qualified by the fact that almost all emotional indicators tap positive instead of negative aspects of intergroup attitudes (e.g., liking and sympathy and not displeasing, anger,
fear, or intergroup threat), and this positive dimension does not seem to be sensitive to intended changes. Indeed, contact-based interventions primarily promote intergroup attitudes by changing negative emotions such as fear or intergroup threat (Pettigrew & Tropp, 2008). Unfortunately, the studies integrated here hardly ever used these measures. Moreover—at least for interventions focusing on modifying negative facets of intergroup attitudes—an increase in sympathy and liking is not necessarily a primary objective. In addition, several intergroup attitude constructs such as tolerance or respect do not inevitably imply sympathy, even if they are not independent from each other (see van Quaquebeke, Henrich, & Eckloff, 2007).

In sum, the finding that the integrated studies neglect negative emotional ratings almost completely as outcome measures is highly unsatisfactory. This is particularly because negative emotions seem to play an important role in the development of serious prejudice and discrimination. For example, intergroup threat and negative feelings toward an out-group are one of the best confirmed risk factors that will lower intergroup contact and impede intergroup friendships (e.g., Riek et al., 2006). Therefore, future intervention programs for children and adolescents should focus more intensively on intergroup emotions, and future evaluation studies should pay more attention to measuring especially the negative part of the emotional dimension of prejudice and intergroup attitudes.

Effect-size differences in outcome assessment have also been ascertained between the categories representing the type of intergroup attitude in which measures of in-group favoritism (e.g., higher rating of in-group members on a positive characteristic such as intelligence) had an effect size close to zero. In contrast, prejudice measures, that is, measures representing some kind of bias on a negative dimension (e.g., being mean), showed higher and significant effect sizes. At first glance, this pattern of results appears to correspond with the central aims of the interventions, because changing negative evaluations seems to be a priority goal of many intervention programs. However, even small differences in in-group favoritism could also have considerable negative effects for out-group members (e.g., Brewer, 1999), although the probability of severe consequences from negative evaluations is perhaps higher. In addition, Brewer (1999) argues that in-group favoritism is close to a more subtle kind of prejudice that may well be harder to change via interventions (see also Dixon, Levine, Reicher, & Durrheim, 2012). Again, future programs should focus more on applying several measures and indicators of intergroup attitudes at the same time in order to investigate effects on divergent dimensions and facets.

Results on program components

This meta-analysis also provided some insights into which programs or program components can possibly be recommended. However, as in other reviews, the high range of intervention programs makes it virtually impossible to state which is the most promising single approach for altering prejudice and improving intergroup attitudes. Because of this diversity, we coded the intervention programs separately on the most highly distinguishing program components and parameters. The advantage of this coding procedure was that we could also enter the diversity of the program components into our meta-regressions in order to rule out confounds. This at least allowed us to examine whether any of these components were superior to others, and it should deliver indications for future developments in prejudice prevention programs.

This detailed analysis revealed that direct contact and training in empathy and perspective taking are the most promising program content components. The contact effect is a well-known result that has been confirmed in several reviews during the last decade (e.g., Aboud et al., 2012; Paluck & Green, 2009; Pettigrew & Tropp, 2006). In addition, our analyses revealed that direct contact leads to better results than any kind of vicarious, extended, or—more generally—indirect contact. However, although the latter shows lower effects, these are still significant. The relatively small effect produced by indirect contact is somewhat surprising in any case, because it has been recommended strongly by some authors (e.g., Wright et al., 1997). However, previous reviews have also reported that some moderating and mediating factors have to be considered when looking at contact effects regardless of whether they are direct or indirect (Pettigrew, Tropp, Wagner, & Christ, 2011). For example, Schofield’s (1995; Schofield & Hausmann, 2004) analysis of school desegregation produced rather mixed findings, revealing that, within this context, contact alone is not a sufficient condition for positive effects. Success seems to depend critically on whether the programs succeed in initiating personal relationships and friendships between members of different ethnic groups (see also Aboud & Levy, 2000; Tausch, Schmid, & Hewstone, 2010). In addition, recent meta-analytic results (e.g., Davies et al., 2011) confirm the central role of personal friendships, and these are impossible or at least difficult to promote in an indirect contact situation. Therefore, program developers should integrate direct contact elements in order to acknowledge the necessary or at least beneficial contribution of the contact situation (see above; Tropp & Prenovost, 2008). However, direct contact is frequently impossible in some social contexts, because members of the out-group are physically not present. In such cases, some kind of indirect contact has to be implemented (e.g., Beelmann et al., 2010; Cameron, Rutland, Brown, & Douch, 2006). Nonetheless, our results indicate the benefits of actively including at least some element of direct contact (e.g., some joint play activities with visiting children or member of the out-group as trainer) that would at least support and complement indirect contact through personal interaction experience.

In addition to the direct contact effects, training in empathy and perspective taking showed high potentials for promoting intergroup attitudes via social-cognitive abilities. The beneficial influence of empathy and perspective taking has long been discussed in the developmental literature on intergroup attitudes (Aboud & Fenwick, 1999; Bigler & Liben, 2007), and a recent meta-analysis has confirmed its importance for prejudice development, especially in elementary school age children (Heinemann & Beelmann, 2011). Although moral or value education was not significant in our meta-regression analysis, our categorical analyses indicate that it seems to deliver some promising results. Promoting moral development is most likely to be confounded with perspective taking and empathy, and any analysis of it may suffer from the relatively small number of relevant training-control comparisons. Within the field of moral development, we clearly need more high-standard evaluation studies, especially because recent studies suggest that it plays a central role in prejudice and intergroup attitude development (Kellen, Kelly, Richardson, & Jampol, 2010; Rutland et al., 2010).

Training in classification skills and reorganizing social categorization has been recommended in a series of studies, especially within the social psychological literature (Gaertner & Dovidio, 2000; Gaertner et al., 2008). However, the results of our meta-analysis fail to confirm these recommendations—at least for children and adolescents. We did not find any significant positive changes, and sometimes even negative effects of training classification skills or restructuring the processing of social categorization. There might be different reasons for this: First, classification skills may not be as important as several authors and social cognitive developmental theories assume. For example, in a meta-analysis on individual prejudice predictors in childhood and adolescence, Heinemann and Beelmann (2011) found that these skills did not correlate significantly with prejudice at any stage of development. Thus, it seems less useful to train these abilities in order to promote intergroup attitudes—possibly because the trainings are restricted mainly to classification skills for nonsocial characteristics.

Second, because social categorization and stereotyping form the psychological basis of intergroup attitudes (Brown, 2010), modifying social categorizations might also have negative effects when training sensitizes members of the target group to social categories that they
were not aware of before the start of the program (i.e., they serve as a root for a subsequent prejudice). Therefore—and in contrast to adult target groups—intervention programs for children and adolescents should probably work on social categorization only in cases in which social categories are already known and used for producing biased social attitudes.

A third possible reason might lie in the type of implementation of social categorization. Several studies have shown that how social categorization takes place seems to be important. For example, Cameron et al. (2006) reported that only reconstruction following the dual identity model (i.e., invoking a superordinate social category but simultaneously retaining the subgroup identity) had positive effects. In contrast, the common in-group model (i.e., summarizing in-group and out-group within the same superordinate social group) and decategorization (i.e., teaching that the corresponding social category is not important) were less effective—at least for the age group under investigation (preschool and elementary school children). In general, results on program content indicate clearly that more research is needed on how to structure social categorization within training programs in order to promote positive intergroup relations in children and adolescents. Simple strategies such as “these categories are not important” or “we all belong to one social group of human beings” do not seem to be enough and are, at least in some cases, even counterproductive.

Besides these substantial considerations, we also identified diverse program parameters as important factors for successful program conduct and implementation. For example, interventions with a trainer who was actively involved in the program administration yielded significantly higher effect sizes than interventions without this element—indeed, the person adopting that role (e.g., teacher, program author, or trained student). This result clearly corresponds to the contact literature in which supervision (or taking an active role) by an authoritative person is assumed to be one of the beneficial conditions during intergroup contact (Tropp & Prenovost, 2008). This interpretation probably also fits in with the surprisingly negative influence of the use of printed materials and group discussions on the mean effect size. Printed materials are frequently used when children teach themselves about prejudice and intergroup attitudes, and group discussions are probably less structured intervention methods. Thus, our findings indicate that training concepts receiving only low support from the teacher or program administrator are less likely to deliver significant effects on improving intergroup attitudes.

The results on content and program parameters clearly have some further implications for the prevention of prejudice and for the promotion of intergroup relations in general. First, future program concepts should maybe focus more on combined and multitheoretical models. Because there are divergent aspects of intergroup attitudes and, apparently, several promising ways to prevent prejudice and to promote intergroup attitudes and relations, it would seem to be worth developing new integrative models of intervention. In addition, we clearly find that multimodal programs have higher and more generalized effects on different outcomes. Therefore, the development of new programs would profit clearly from a systematic combination of different intergroup approaches. At the same time, however, it should not neglect the most effective elements such as structured contact experiences and training in perspective taking. For example, Beelmann et al. (2010) combined contact experiences with training in different sociocognitive skills (including perspective taking) and multicultural information and found relatively high long-term effects on different measures of intergroup attitudes. Nonetheless, these approaches mark perhaps only the beginning of a new era characterized by a more comprehensive approach to prejudice prevention research combining multiple theories, methods, contents, and outcome assessments.

Second, because an active trainer is apparently a crucial factor in promoting intergroup attitudes, intervention concepts should not be restricted to didactic materials alone. Programs need to integrate an interaction component involving adults or competent partners. At the same time, we could assume (even though this was only a statistical trend) that interventions should be integrated into daily school activities, because implementing a program within the curriculum seems to be favorable. In sum, findings on program parameters indicated that interactive, structured, well-planned, and well-implemented programs seem to be more valuable and successful—an outcome that has been confirmed frequently in prevention research with children and adolescents in general (Durlak & DuPre, 2008; Durlak et al., 2011).

Results on different target groups

In addition to the results on content and program parameters, we identified two further variables that correlated systematically with effect sizes. On the one hand, higher effects were found when the target out-group consisted of persons who are handicapped compared to ethnic out-groups. The fact that attitudes toward the disabled are apparently easier to improve is probably due to several factors. First, although integrated schooling has been implemented broadly in several countries, contact with the handicapped is perhaps less probable in everyday life than that with ethnic groups—at least in multietnic countries such as the United States. Moreover, the majority of studies addressing intergroup attitudes toward the handicapped were performed outside integrative schools. As a result, children may have less experience in interacting with children who are disabled and thus have more potential to learn something about this group and promote their attitudes toward them—at least in information-based and contact-based interventions. Although we have no comparable data on the preintervention level of prejudice, it is a well-known result in intervention research that groups showing higher problems or less competence profit more from those interventions than less burdened groups (Beelmann & Raabe, 2009; Durlak et al., 2011) simply due to ceiling or floor effects in outcome measures.

Second, the intergroup threat is probably higher in intergroup contexts with ethnic out-groups compared to intergroup contexts with people who are disabled, especially in high conflict areas or multietnic countries. This may hold for elderly people as a target group as well, for whom we also found relatively high effects. This higher mean level of intergroup threat may well be responsible for the significantly lower effect sizes in ethnic prejudice and intergroup attitudes, because of more difficulties in concrete implementation of programs. However, we found virtually no intervention studies addressing this issue.

Finally, western societies may well have stronger antiprejudice and antidiscrimination norms for the handicapped and for elderly people than for ethnic minorities, probably again because of different level of intergroup threat. Therefore, program goals with these out-groups are either easier to attain or less sensitive to interfering norms in the social context. However, despite this relatively unfavorable result for ethnic out-groups, it should not be forgotten that there are also significant effect sizes for improving intergroup attitudes toward other ethnic groups. Future research should focus more on the concrete social context and the ethnic conflicts underlying attitudes in order to optimize intervention efforts in this area.

The second variable showing a systematic correlation with effect sizes was the in-group participating in the intervention. Studies with minority groups (e.g., in which they are the in-group and the intervention addresses prevention of prejudice against majority children) revealed nonsignificant effect sizes and were far less effective than interventions with majority groups (e.g., addressing prejudice against minority children). One reason for this result is probably the poor fit between interventions and the development of intergroup attitudes in minority groups (see also Aboud et al., 2012). A recent meta-analysis has shown that the developmental course of intergroup attitudes differs for minority compared to majority groups (Raabe & Beelmann, 2011).
Majority (high-status) children showed an increase in the level of prejudice against minority children between ages 3 and 5 to 7 and a subsequent decrease up to ages 8 to 10. Minority (low-status) children, in contrast, start to increase their prejudice against majority children at a later age and fail to show the decrease found in majority children. These different developmental trajectories are probably due to different developmental processes, indicating the need for different kinds of interventions at different developmental stages. For example, the decrease in majority children could probably be ascribed to emerging sociocognitive abilities (e.g., Aboud, 2008; Bigler & Liben, 2007). If so, this indicates the benefits of training in these abilities at this age to promote natural development. In minority children in contrast, the later increase probably relates to specific identification processes (e.g., after experiencing first discrimination through majority children). This indicates the benefits of more training in self-efficacy, coping with negative experiences, or establishing specific contact opportunities to majority members (e.g., Banaji, Baron, Dunham, & Olson, 2008). Therefore, developmental knowledge on different developmental courses should lead to different prevention concepts. However, the four studies addressing minority children (Aboud & Fenwick, 1999; Avci-Werning, 2004; Feinman, 1982; Walker, 1971) used essentially the same intervention as that used for majority children.

In addition, although concrete intervention planning should depend on the specific in-group–out-group constellation in a given social environment, the aforementioned developmental knowledge leads us to assume that minority groups should not simply be integrated within programs for majority children and adolescents. For example, information on and contact to the out-group is perhaps redundant for minority children, because they are exposed to this out-group constantly in their everyday lives. Therefore, it seems advisable to apply different antibias interventions for majority and minority members. However, the present comparison of majority versus minority children and adolescents and the conclusions made are based on a very small number of studies for minorities, indicating a need for further research to derive more valid conclusions on this important topic.

One final word should address the age of the participants as effect moderator. We failed to find differences in effectiveness of programs according to the age group treated. On the one hand, this is surprising, because different phases of intergroup attitude development would lead us to expect age-related differences in the amenability to intervention programs (Raabe & Beelmann, 2011). In addition, prevention research is largely based on the assumption that “earlier is better,” suggesting that prevention measures should be more effective in younger age groups.

On the other hand, there may be several reasons why age could not be identified as a significant moderator. First, although nonsignificant, we nevertheless found a slightly higher effect size for the age group between 8 and 10 years, which corresponds to highly distinctive developmental changes—at least for majority children (see above). Therefore, results may indicate small age-related differences that could not be confirmed because of low statistical power. Second, age is normally confounded with other variables (e.g., intervention concept, level of prejudice). As in many meta-analyses, these confounds are mostly neither easy to control (e.g., in order to consider the level of prejudice, we need comparable pretest scores) nor easy to take into account systematically when interpreting results. For example, we found no systematic relation between the content and program parameters of the intervention and age group apart from a higher percentage of socio-cognitive training programs in older children from age 8 on. This confound is probably responsible for some bias when comparing the results for age groups, but cannot explain entirely why they remain largely comparable over the course of development. Therefore, the lack of age-specific results suggests that even effective prevention of prejudice and promotion of positive intergroup attitudes is possible at all ages as long as the underlying intervention concepts are developmentally appropriate (see also Aboud et al., 2012).

Limitations and future research on promoting intergroup relations

This meta-analysis also has its limitations that simultaneously deliver indications for further research on promoting intergroup attitudes and relations. First, the great diversity of intervention programs and approaches makes it almost impossible to identify clearly defined programs that are more effective and therefore to be recommended. In addition, we can make no systematic statement about the appropriateness of the programs within their specific social context and how they could be tailored to these contextual requirements. In general, programs seem to have clear deficits in relation to developmental theories and existing empirical knowledge on intergroup relations. For example, only a few intervention concepts analyzed in this meta-analysis derive their content and further program parameters such as age at start from modern developmental theories on intergroup attitudes (e.g., Bigler & Liben, 2007; Nesdale, 2004). In addition, although empirical research has revealed that several risk factors are involved in prejudice development, only a few programs have multicomponent concepts (e.g., Beelmann et al., 2010). However, this recommendation for closer links between developmental research and intervention concepts also indicates the need for further developmental studies. For example, up to now, it is relatively unclear how stable intergroup attitudes actually are in childhood and adolescence, and which prejudice thresholds deliver a clear negative developmental prognosis and are linked to negative intergroup behavior in the future. A few longitudinal results indicate some strong continuities over the course of development as well as a linkage to behavior problems (see Beelmann & Heinemann, 2011; Raabe & Beelmann, 2011). This is important information for constructing intervention programs and deciding at what age they should be applied. Therefore, future social psychological and developmental research should focus more on applying longitudinal designs and developing closer links to prevention and intervention programs in this field. This, in turn, should make these interventions more effective (see also Beelmann, 2011; Killen et al., 2011).

Second, in light of the only moderate effect sizes found here, it seems advisable to look for alternative intervention approaches in the future such as programs with parents, teachers, and peers in order to promote their function in socializing toward more positive social norms (see Killen et al., 2011). This strand of intervention research is quite neglected despite the strong ties between, for example, parent and child intergroup attitudes (Degner & Dalege, 2013).

Third, most of the systematic intervention research has been performed in North America and Europe. However, problems in intergroup relations are universal phenomenon that needs to be addressed in many countries—especially in those with massive conflicts between ethnicities or other social groups. The few studies coming from Israel (e.g., Slone, Tarrasch, & Hallis, 2000) have taught us that principles of intergroup contact and promoting empathy and perspective taking may also hold for these social contexts. However, especially in the context of strong and long-lasting historical conflicts such as those in the Middle East or some regions of Africa, it is in some way questionable whether these kinds of psychological programs for children and youth alone will suffice. Alternatively, we should think about more multidimensional and systemic interventions including, for example, new legislation, peace education within society, and reducing intergroup threat by promoting reconciliation between conflict parties (Bar-Tal & Rosen, 2009; Staub, 2006). Research on this integrative approach to intergroup conflicts is only just beginning (Nadler, Malloy, & Fisher, 2008; Tropp, 2012), but is much needed to apply existing psychological knowledge to solve problems in these adverse social contexts.

Fourth, turning to the type of outcome assessment, only a small variety of different outcome measures was used, and these do not represent the multifaceted character of intergroup attitudes. Most studies used cognitive intergroup attitude indicators to measure prejudice and applied questionnaires to assess it. More data are needed, especially to estimate the effect on the negative emotional facets of intergroup
benefit rather than suffer from social diversity (Dixon et al., 2012; Tropp & Mallett, 2011).

References


