Developmental Review 32 (2012) 307–336

Contents lists available at SciVerse ScienceDirect

Developmental Review

journal homepage: www.elsevier.com/locate/dr

Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review

Frances E. Aboud [a,*], Colin Tredoux [b], Linda R. Tropp [c], Christia Spears Brown [d], Ulrike Niens [e], Noraini M. Noor [f], the Una Global Evaluation Group

[a] McGill University, Canada
[b] University of Cape Town, South Africa
[c] University of Massachusetts Amherst, USA
[d] University of Kentucky, USA
[e] Queens University Belfast, Northern Ireland
[f] International Islamic University Malaysia, Malaysia

ARTICLE INFO

Article history:
Received 29 September 2011
Revised 14 May 2012
Available online 10 September 2012

Keywords:
Prejudice reduction
Attitude change
Ethnic bias
Multicultural education
Intergroup relations
Children
Intervention
Program evaluation

ABSTRACT

A systematic review was conducted of studies evaluating the effects of interventions aimed at reducing ethnic prejudice and discrimination in young children. Articles published between 1980 and 2010 and including children of 8 years and under were identified, harvested, and assessed for quality, both for the exposure/program as well as for the evaluation. In total, 32 studies (14 contact and 18 media or instruction) yielded 62 effects on attitudes and 59 effects on peer relations. An overall count of the positive (40%), non-significant (50%), and negative effects (10%) indicate a mixed picture. Overall, more attitude effects (55%) than peer relations effects (25%) were positive, and media/instruction (47%) was more successful than contact (36%). Most of the effects were observed with children from a majority ethnicity: 67% of the attitude effects were positive, and media/instruction and contact were equally effective at delivering these. Few differences were found as a function of the quality of the exposure and evaluation, but differences were found depending on the context of exposure (naturally occurring or experimental manipulation) and research design (random assignment or self-selection). In conclusion, the findings were more mixed than expected, though sufficiently strong studies exist to provide lessons for future research.

* Corresponding author.
E-mail address: frances.aboud@mcgill.ca (F.E. Aboud).

0273-2297
http://dx.doi.org/10.1016/j.dr.2012.05.001

Introduction

Once it was realized that prejudice developed at a very young age in children, psychologists and educators increased their efforts to examine its development and possible reduction. Descriptive and experimental research has by now mapped out the changes that occur with age and the factors such as majority/minority status that influence changes (see reviews, e.g. Raabe & Beelmann, 2011). It is clear that in many parts of the world, with populations of multi-ethnic background, prejudice begins around 4- to 5-years of age. Prejudice, in most scholarly work, is defined essentially as negative evaluations, beliefs, or feelings directed at people because of their ethnicity (Brown, 1995). The behavioral component, called discrimination, entails treating others differently because of their ethnicity, such as name-calling and social exclusion. There may be different age trajectories for prejudice and discrimination, but both are considered detrimental to social harmony and productivity, and so worthy of attention. Moving beyond the descriptive research, we reviewed studies that evaluated interventions to reduce prejudice and discrimination. To avoid assuming that young children are prejudiced or that the goal is narrowly to reduce prejudice, we took a broader view of our objective, namely to examine interventions that enhance respect and inclusion. We viewed *respect* as the positive attitudinal goal and *inclusion* as the positive behavioral goal of interventions.

The focus on children has emerged as a result of our understanding that youth sets the stage for future attitudes and behavior. Prejudice, particularly toward visible minorities, is now known to begin in early childhood between 3 and 6 years of age (Raabe & Beelmann, 2011). There is still debate as to whether programs should be targeted at an early age when prejudice is quickly developing or in middle childhood when prejudice diverges due to environmental input (Raabe & Beelmann, 2011). The case for early intervention rests on the finding that prejudice is self-perpetuating because prejudiced children avoid disconfirming experiences and information. For example, children with cross-ethnic friends in the early grades are more likely than those without such friends to have an integrated social network as adolescents and adults (Ellison & Powers, 1994; Patchen, 1982). Contact with other ethnic groups has increased as a result of migration and the reduction of other constraints such as apartheid and economic inequality. This has brought changes to once homogeneous neighborhoods. As communities and schools become more socially inclusive, children are being provided increased opportunities to befriend those from different ethnic backgrounds. Educators and policy makers are learning from past experience as they devise programs to prevent or reduce prejudice and discrimination (see Banks, 2009 edited book on multicultural education). Increasingly, these programs are focused on the early childhood years (3–8 years) to provide children with the social and cognitive skills needed to work and play with peers (Nadeau, Kataoka, Valerio, Neuman, & Elder, 2011).

Current theories of prejudice development tend to include both sociocontextual and sociocognitive constructs (e.g., Bigler & Liben, 2006). Although theories underlying programs to *change* prejudice build on these same constructs, they tend to focus more on changes to the context, such as contact and instruction, that may in turn influence cognitive and emotional processes. However, they rarely address age-related cognitive structures such as egocentrism. It is reasonable to expect that theories of change will differ from theories of natural development, but the neglect of age-related processes is an oversight that might jeopardize success.

At one end of the sociocontextual continuum is the theory of contact developed by Allport (1954) that outlines specific conditions under which intergroup contact reduces prejudice, such as equal status and authority support. Kenworthy, Turner, Hewstone, and Voci (2005) proposed a number of psychological processes that mediate contact and respect, such as reduced stereotyping due to attention to individual rather than racial differences, reduced anxiety, feelings of closeness to the other such as empathy and perspective-taking, and recognizing social norms of contact. Meta-analyses of the relation between contact and attitudes demonstrate that all conditions may not be necessary for contact to be effective (Pettigrew & Tropp, 2006), but that friendship, the gold standard of contact, is strongly associated with respectful attitudes (Davies, Tropp, Aron, Pettigrew, & Wright, 2011). Programs that allow for and intentionally generate opportunities for close contact are informed by this theoretical framework.

Alternative theories of prejudice reduction focus on direct exposure to information that changes the way people think and feel about other groups. In this case, the context is intentionally manipulated, for example, through media and instruction. These theories are useful when contact is minimal. For example, Bigler and Liben (2006) outline antibias information that is not only associated with respect but is capable of breaking down negative generalizations. Repeated exposure to inclusive ways of organizing a class of students, repeated reinforcement of thinking about multiple ways of classifying people, and arousal of anger at others' experiences with unfair discrimination are some examples. In addition to antibias information, there may be training in skills to confront bias and in emotions such as empathy. The psychological processes being targeted are similar to those expected to change as a result of contact, but the strategy is communication. Constructs relevant to communication theory are therefore included, namely the antibias message, the sender of the message (e.g., peer or teacher, ingroup or outgroup), and the age and identity of the recipient (e.g., see Aboud, 2005; Petty & Cacioppo, 1986; Slater & Rouner, 2002).

The use of theory in developing interventions to reduce prejudice is now more common as a result of well-formulated theories being published, including the seminal work of Katz and Zalk (1978). Consequently, interventions tend to explicitly mention contact theory or antibias theory. Interventions using media may use both theories, by presenting information about two cross-ethnic friends who capture the conditions raised by Allport, along with antibias information. It is also conceivable that contact could be combined with specific training on reconciling differences and multiple classification to reduce egocentrism and dichotomous thinking.

Recent reviews have drawn attention to the fact that while the gap from theory to research is narrowing, the gap from research to practice is still wide (Paluck & Green, 2009). For example, conditions known to enhance the benefits of contact are not always implemented in community and school settings. Furthermore, the rigorous designs used in small-scale laboratory research are rarely found in evaluations of school programs. A more serious concern is that reviews to date have not systematically integrated studies of young children that examine the outcomes of interventions to reduce prejudice and discrimination. Instead, they have taken a broad age range that ignores most research on children (Paluck & Green, 2009). More importantly, they tended to summarize successful interventions (e.g., Aboud & Levy, 2000; Cameron & Turner, 2010, chap. 14; Pfeifer, Brown, & Juvonen, 2007) that may not adequately represent the variety of findings. In order to focus future efforts on the promising strategies and eliminate ineffective ones, a systematic review of interventions for young children is needed. It would also draw attention to design and measurement issues that could strengthen evaluations.

Therefore our objective was to examine studies on the effectiveness of various exposures that reduced prejudice and discrimination or increased respect and inclusion. Our focus was on children in the early childhood age range of 8 years and under. We included studies using this age range where older children might also be included; where possible, disaggregated data for the desired age range was extracted.

Method

Guidelines produced by the Evidence for Policy and Practice Information and Coordinating Centre for systematic reviews were used to identify, screen and describe all relevant studies (2007, retrieved January 2009 from eppi.ioe.ac.uk/cms/).

Study Selection

The search strategy included locating relevant articles in electronic databases published from 1980 to August 2010, from PsychINFO, ERIC, Current Contents and several others. This was done over three successive years during the month of July (2008, 2009, 2010). Keywords used initially were*: prejudice reduction, anti-bias, multicultural curriculum, empathy training, intergroup training, diversity training, jigsaw classroom, racial bias*; the search was narrowed by combining with *child* or *children*. Because the search did not turn up articles that we knew existed, we expanded keywords used to search the databases to include: *prejudice, inclusion, exclusion, ethnic bias, ethnicity + attitude, attitude change, peer relations* or *friendship + ethnic* or *cross-ethnic* or *interethnic* or *interracial, sesame*. Keywords such as *intervention* and *program evaluation* were added to the former set.

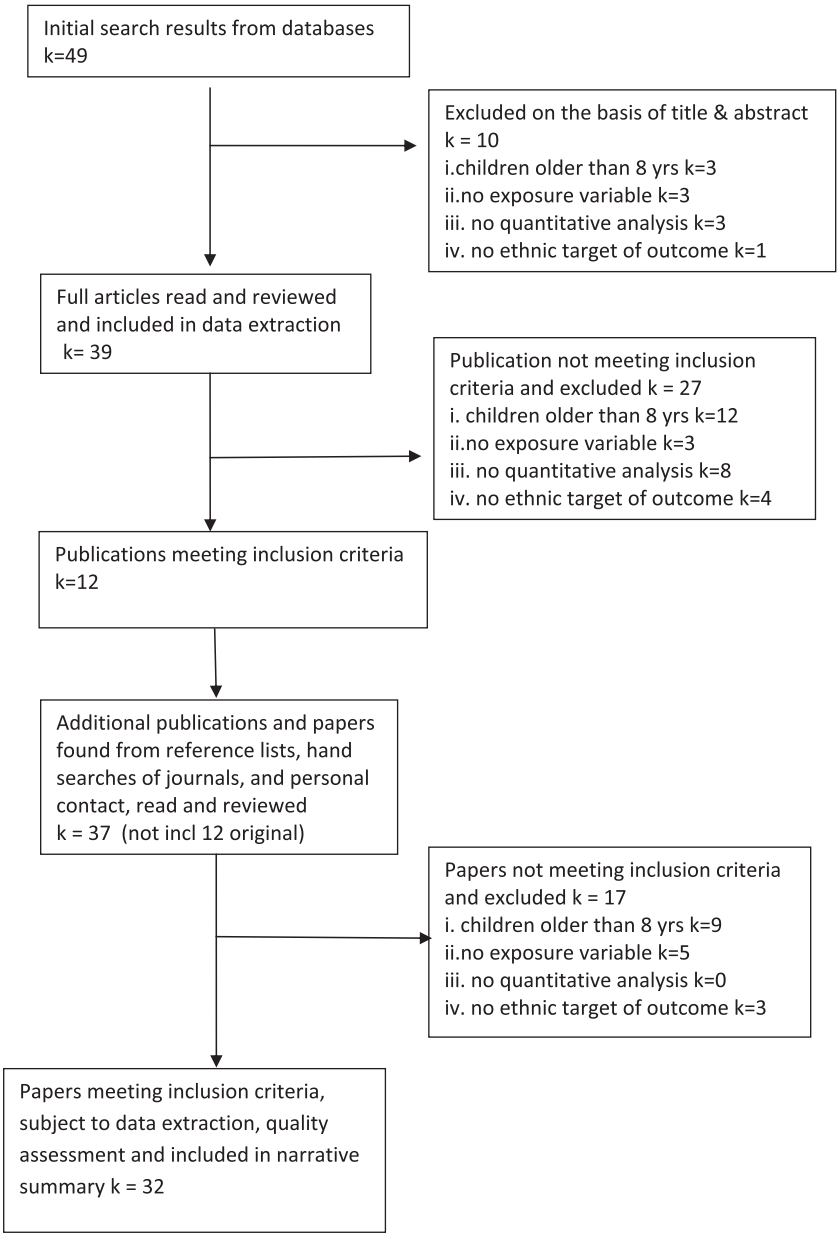

Fig. 1. Search results and publication selection procedures.

Additional titles were retrieved by manually searching the reference lists of review articles published since 1995 and of all included full-text articles. We also conducted hand searches of the following journals from 2000 to mid-2010: Child Development, Developmental Psychology, International Journal of Behavioral Development, International Journal of Intergroup Relations, Journal of Social Issues, and Educational Psychology. Personal communication with authors known to conduct child-related research on prejudice up to 2011 and authors identified in Dissertation Abstracts elicited several unpublished papers.

Inclusion criteria

An article was included for a full screening if it met all of the following criteria: (i) it included children of 8 years and under; (ii) it included one or more of the key outcome measures, namely attitudes (e.g., evaluation, respect), personal relations (e.g., liking, friendship, inclusion, exclusion, or intention to interact), and social behavior (e.g., voluntary play, dyadic communication, sharing, helping, cooperating, verbally responding to name-calling); (iii) an independent exposure variable, either naturally occurring or intentionally implemented, was examined for its association with one of the outcomes; (iv) original data were statistically analyzed for the association between the exposure and the outcome; (v) the exposure and the outcome concerned naturally existing ethnic, racial, religious, national or language groups. A number of studies were excluded either because they assessed attitudes toward novel (experimenter-created) groups (e.g., Patterson & Bigler, 2006), or because the exposure entailed presentation of only a small number of stereotypic attributes intended to arouse prejudice (e.g., Abrams, Rutland, & Cameron, 2003, presented English children with stereotypic fans of the German football team). Published and unpublished research was included, and although we could review articles in French and Spanish as well as English, none of the former was found. We did not exclude articles on the basis of quality.

For the purpose of our review, prejudice was defined as negative evaluations of people on the basis of their group membership and respect was defined as positive evaluations. They were usually assessed in terms of positive and negative evaluative terms such as 'friendly' and 'mean', but not non-evaluative stereotypes such as 'shy' and 'musical'. Inclusive vs. exclusive personal relations were defined as peer relationships described as positive or negative and assessed through self-report of befriending or avoiding/conflicting during peer relations, friend nomination of a peer, reported self-disclosure within a peer relationship, or reported intention to interact. Social behavior seen as supportive or discriminatory was defined as overt actions that included or avoided/criticized another person. It was usually measured through observation of talking, playing and working with a peer, and standing up for a peer who was called names.

As illustrated in Fig. 1, the initial search generated 49 empirical studies and 52 reviews, books and program manuals. Titles and abstracts were read by 3 raters. Only the former 49 included empirical studies and so were relevant for the review. Of these, ten were excluded because all the children were over 8 years (3), or the study had no exposure variable (3), or there was no quantitative data analysis (3), or it was not about ethnicity (1). Thirty-nine articles were fully read by two or more raters and 27 more were excluded, leaving 12. A search of references, recent journals, and personal contacts yielded 37 additional articles of which 17 were excluded. Consequently, the 12 + 20 were abstracted in order to describe information relevant, among others, to the exposure variable and the outcome measure. Consequently, 32 studies are included in this review. Because of the heterogeneity of exposures, designs, and outcomes, relative to the number of studies, and the few randomized controlled designs, a meta-analysis may be of little value, and the conclusions reached there of questionable validity.[1] We therefore conducted a systematic narrative review, in preference. Exposure variables included three main categories: contact with another ethnic group ($k = 14$), media exposure to another ethnic group

[1] We debated at some length whether to do a meta-analysis on the studies we identified for review, but on reflection it seemed clearly inappropriate. Consider the following caveat about conducting meta-analyses from Lipsey and Wilson's (2001) book on the topic: "... it is not generally appropriate to combine study findings derived from different research designs and appearing in different statistical forms, even if they deal with the same topic. For instance, experimental studies of treatment for depression using treatment vs. control group comparisons generally would not be combined with observational studies in which level of depression was correlated with level of service received. Though both types of study deal in some fashion with the relationship between treatment and depression, the differences in the research designs, the nature of the quantitative relationships constituting the findings, and the meaning of those findings is so great that they would be difficult to incorporate in the same analysis." (p 3). A similar argument is made in the Cochrane handbook of systematic reviews: "...If studies are clinically diverse then a meta-analysis may be meaningless, and genuine differences in effects may be obscured. A particularly important type of diversity is in the comparisons being made by the primary studies. Often it is nonsensical to combine all included studies in a single meta-analysis: sometimes there is a mix of comparisons of different treatments with different comparators.... it is important not to combine outcomes that are too diverse" (http://www.mrc-bsu.cam.ac.uk/cochrane/handbook, Chapter 9). In the present case we are reviewing studies that used group comparisons, with and without randomization, studies that are post hoc and/or correlational in nature – and these are just a few of the methodological variations. The use of meta-analysis as a review method in the face of this great heterogeneity seems ill-advised, if not simply proscribed.

(k = 9), and exposure to multicultural or antibias instruction, which may include empathy or anti-bullying training (k = 9).

Data extraction

Detailed information about each study was recorded in order to identify characteristics of the intervention (i.e., the exposure variable) and its evaluation.

Data on the following exposure variables were coded and entered into a spreadsheet: author, publication date, ethnic group to whom exposure was directed, age of child participants, the analytic level at which the exposure was targeted (e.g., whole school, individual), who/what delivered the exposure (e.g., media, teacher, parent), nature of the exposure variable (e.g., video, diverse student body), format of the exposure variable (e.g., instruction, daily direct contact, close friendship, video of exclusion, tapes or stories of name-calling), duration of exposure (e.g., at least 1 year, 24 times over 8 weeks, 5 min session).

Data on the following evaluation assessment variables were coded and entered into a spreadsheet: overall design (e.g., correlation of naturally occurring variation in exposure, experimental analogue with control over exposure, exposure developed as a program with less control over its implementation) and design properties (e.g., cross-sectional, longitudinal, intervention-control, pre-post test or post-test only, randomized or non-randomized to exposure). Measures or manipulation of exposure (e.g., contact, media, instruction) and measures of child outcomes were outlined and the association was recorded as a positive, negative, or non-significant relation. Outcomes were divided into attitudes and peer relations/behavior. Mediators and moderators, such as ethnicity and age of children, were noted in the comments column. At least two raters extracted the data for each study.

Quality assessment

Two assessments of quality were undertaken: quality of the exposure, e.g., how age-appropriate and theory-informed was the intervention; and quality of the evaluation, e.g., how sound were the measures and design (see Table 1). Each study was rated strong, medium, or weak on a number of qualities (e.g., van Sluijs, McMinn, & Griffin, 2009).

The quality of the exposure variable was scored according to nine criteria: strong exposures were those where activities/materials were linked to a theory and age-appropriate, there were 5 or more sessions/hours, it was observed to be implemented in a way that was faithful to the intended exposure, the implementer was someone normally in a position to conduct such a program when implemented on a large scale, and there was some formative research or piloting of the program to ensure its viability. In some cases, the exposure variable was simply contact with a diverse student body and in other cases it was a fully-fledged program delivered to children at a school. The qualities were taken from Davidson et al. (2003), who suggested that minimal intervention (exposure) detail should be described in research reports in order to know what children were exposed to and how it compared with control-group exposures.

The nine qualities of evaluation were taken from Deeks et al. (2003) and Downs and Black (1998) to address internal and external validity. They included strength of the design, sample, measures, statistical analyses, and time of assessment. High marks were given if a study used a conventional measure and provided reliability estimates, used a randomized design, with a relatively heterogeneous and sufficiently large sample, research assistants were blind to grouping, confounders were identified, analyses were not overly numerous or if so used a correction and adjusted for clustering, and there was a follow-up assessment.

Inter-rater reliability was conducted on 75% of the studies. Kappa coefficients were calculated for each of the qualities: the medians of the coefficients were .73 for quality of the exposures and .62 for quality of evaluation. Agreement was actually higher than these values reflect because kappas could not be calculated for several qualities where there was insufficient spread across the categories, i.e., because 90–100% of the scores were agreed to be in one category (e.g., strong, or weak). Discrepancies were then resolved through discussion, and the remaining 25% of studies were scored through consensus.

Table 1
Quality criteria used to assess exposures (e.g., programs) and evaluations.

Quality of program	Score	Quality of assessment	Score
Theoretically grounded		**Data collection methods/measures**	
S = explicit theory about mechanism of change or development	S = refers to and is developed from an explicit theory of how exposure → outcome	**S** = conventional measure with reliability	**S** = not homemade; own alpha calculated
M = mechanisms of change not articulated but could be linked to theory		**M** = conventional measure OR reported reliable	Validity of measure is not included
W = no clear theoretical assumptions		**W** = measure not conventional, no reported reliability	
Activities or materials linked to theory		**Study design concerns main exposure var**	
S = activities and/or materials clearly derived from or linked to theory	Exposure activities or materials reveal link to theory	**S** = randomized at individual or group level	S = Ss assigned to gp
M = implicit but could be linked to theory		**M** = two group non-randomized design with pre-post or post-only	M = self-selected levels of contact or friendship
W = no link to theory		**W** = Intervention group only, pre-post	
Duration of intervention		**Selection Bias**	
S = 5 or more sessions or 5 or more hours		**S** = good heterogeneity in socio-demo variables	
M = 2–4 sessions or 2–4 h		**M** = some heterogeneity in socio-demo variables	W = atypical group, e.g. private school
W = 1 session		**W** = very little heterogeneity, e.g. only 1 school	
General useable applicability of program		**Unit of Analysis**	
S = applicable to most or all children	S = to at least one majority and one minority	**S** = individual or individuals in class adjusting for clustering if delivered to a group	S = if exposure is to indiv, or if exposure is to group & there is expected intracluster corr adjusted for
M = applicable to only one ethnicity		**M** = individual but not adjusted for clustering	
W = applicable to only prejudiced or victimized children		**W** = group-level only such as class	
Age-appropriate		**Blinding**	
S = materials & topics appropriate for age		**S** = RA were blind to research question & group assignment	S = RAs could not have known about research question or group assignment
M = designed for older but tried with younger		**M** = RA were blind to group assignment but not research question	

(*continued on next page*)

Table 1 (*continued*)

Quality of program	Score	Quality of assessment	Score
W = not considerate of age-appropr		**W** = RA were not blind to group assignment, or it was difficult to determine	
Fidelity		**Confounders**	
S = some observation of fidelity or action research or manipulation check	S = exposure is observed or a convergent measure used to confirm exposure as intended, e.g. for contact this means observe that contact or friendship is occurring	**S** = groups tested as comparable at baseline or the differences controlled in design or analysis, i.e. diff in DV at baseline and in other relevant var	S = may be controlled in random assignment or repeated design, or in covariate analysis. If exposure is self-selected, confounds are identified
M = manual used but no monitoring of intervention		**M** = only some confounders were reported and controlled in design or analysis BUT not all.	
W = no clear guideline for intervention and no monitoring		**W** = few or no confounders were reported and controlled for in the design or analysis	
Transferability to other Implementers		**Sample Size totaled across consistent findings**	
S = same as in real world e.g. teacher		**S** = if 80 or more per group	S = across all studies in an article if they converge in findings
M = trained assistant, e.g. employee of researcher		**M** = if 50–80 per group	
W = researcher		**W** = less than 50 per group	
Child-friendly setting		**Analysis of Outcomes used in our review**	
S = delivered in school, community, home		**S** = less than 5 or similar outcomes combined or Bonferroni adjusted (minimize Type 1 error)	S = Outcomes used for our purposes are few or significantly overlap, or adjusted
M = not a child-setting, but neutral		**M** = 5–9 analyzed separately with no adjustment	
W = child unfriendly setting		**W** = 10 or more analyzed separately with no adjustment number of hypotheses tested	
Formative research or piloting of prog	S = some preliminary work to see if activity/materials are suitable, effective for that contextM = others have done this in previously publ work	**Participants post-test delay**	
S = prior assessment of materials, activity		**S** = immed post + delayed follow-up 30+days	S = If exposure stops and child tested after 1 month
M = based on prior expt'l research		**M** = delayed posttest – within 30 days	
W = done cold		**W** = only immediate post assessment or assessment while exposure is ongoing	

Results

Research characteristics

The reported characteristics of the 32 studies are given in Table 2. Indicative of the current move toward examining ways of increasing respect and inclusion, 18 of the 32 studies were published/presented from 2006 onwards. A meaningful division of exposure into two sufficiently large categories was made, namely direct cross-ethnic contact ($k = 14$ studies) and media/instructional (18 studies), the latter combining media contact ($k = 9$ studies), and materials/instruction regarding culture and antibias messages or programs ($k = 9$ studies). The interventions in each category were similar in terms of their theoretical underpinnings and exposure variables. Most of the contact studies were with school-aged children of 6–8 or more years, whereas many using media/instruction input were conducted with preschoolers aged 3–5 years.

The nature of exposure was controlled by researchers to varying degrees: highly controlled through experimental manipulations conceived as analogues of what might eventually be a program ($k = 11$), much less controlled when delivered as a planned program implemented by a teacher or parent ($k = 13$), or entirely natural with no control or plan regarding the exposure ($k = 8$). The manner in which children received the exposure variable was considered an important part of the design: children were randomly assigned to the exposure, or were part of an intact group that usually received a certain exposure non-randomly, or children individually self-selected their level of exposure as when they selected friends. The impact of the exposure variable was studied using different time lines: self-selected exposures were often studied in a cross-sectional or longitudinal correlational analysis, whereas randomized and non-randomized groups were compared using pre-post or post-only assessments of the outcome.

Quality of exposures/programs

Table 2 presents the breakdown of studies according to the quality of the exposure, namely whether the contact or media/instruction regarding other ethnic groups used best practices. To reduce the nine quality ratings to a single score, we gave strong ratings a score of 3, medium 2 and weak 1, and then summed the scores for each study. The overall mean quality was 22.53, ranging from 17 to 26 out of a possible 27. Contact studies were a little stronger in that they took place in a natural, child-friendly setting, over a long duration, and were clearly theory-based. Contact studies suffered from having no manipulation check on whether contact actually took place at the individual level. Media and instructional exposures had weaker programs because they were often short in duration, had no manipulation check on whether children received the intended message, and no formative or pilot work to help in developing the exposure for this group of children.

Quality of evaluations/assessments

The quality of methods used to evaluate the effectiveness of the exposure was similarly scored. The overall mean quality of evaluation was 18.34, ranging from 15 to 23 out of a possible 27. Many evaluations of contact, media and instructional interventions suffered from using unconventional measures with reported reliability but no validity, and with a small convenience sample that appeared to be homogeneous. Designs as expected were usually a two-group non-randomized arrangement, and most assessed the outcome immediately after the exposure and sometimes within 30 days but not later. Administration of the exposure to groups of children was rarely followed with an analysis that adjusted for intra-cluster correlation. In many cases, our kappa coefficients for quality assessments were lowered as the result of lack of clear information in the article about important criteria, such as blinding of data collectors and statistical handling of confounders.

Because the 32 studies used a variety of methods, ranging from randomized group experiments to observations of interactions, a meta-analysis was not appropriate (see footnote 1 for further clarification). So we supplemented our narrative review with a frequency count of effects that were positive,

Table 2
Characteristics of the included research.

Characteristic	Contact studies (k = 14)	Media/instructional (k = 18)
Date of publication		
2006–2011	8	10
2001–2005	3	5
1996–2000	1	3
1991–1995	0	0
1980–1990	2	0
Ages of children		
8 years and older	4	3
6–7 years (and possibly older)	5	4
4–5 years (and possibly older)	3	4
3 (and possibly higher)	2	7
Exposure control		
Naturally occurring	8	0
Experimental manipulation	1	10
Planned, less controlled program	5	8
Design		
Self-selected exposure	6	0
Randomized into 2+ groups	1	11
Intact non-randomized	7	7
Outcome variable		
Evaluative attributes	11	12
Social distance (liking)	2	7
Actual Friendship	4	3
Social behaviors	3	2
Quality of the Programs (score)		
25–26	3	2
23–24	9	4
17–22	2	12
Quality of the Evaluation (score)		
20–23	4	4
18–19	3	8
15–17	7	6

non-significant, or negative. Effects were also examined in light of important quality criteria, such as sample size and design.

Findings

The findings and quality of each study are provided in the Appendix. Additional tables summarize the findings across all studies by tallying the number of effects reported for attitudes and for peer relations/behavior. Several effects may be reported from one study. The outcomes of contact exposure on attitudes (k = 35 effects) and on peer relations/behavior (k = 35 effects), and the outcomes of media/instruction exposure on attitudes (k = 27 effects) and on peer relations/behavior (k = 24) effects are presented in Table 3. An overall count of the positive (40%), non-significant (50%), and negative effects (10%) indicates a mixed picture. Few outcomes were negatively affected by exposure, but positive outcomes were not in the majority; however, a binomial test showed that the number of positive outcomes was higher than expected by chance variation ($p < .04$). There was somewhat more support for effects on attitudes where 55% of the effects were positive (binomial $p < .001$) than for peer relations/behavior where only 25% were positive (binomial $p > .86$). There was also more support for media/instruction where 47% of overall effects were positive (binomial $p < .02$) than for contact where 36% were positive (binomial $p > .28$).

Our tally of overall effects appears to be unrelated to sample size. We examined whether, for example, attitude outcomes were more positively influenced than peer relations because the former had larger sample sizes. This was not the case: mean sample sizes for attitude effects were 119.4 and

136.0 for peer relation effects. Likewise, mean sample sizes for media/instruction were not larger than for contact (M = 120.3 vs. 132.5, respectively). The only aspect of quality that determined findings was the design, namely how participants were assigned to condition. These findings are now described.

Findings are organized in Table 4 according to the research design: random assignment, intact groups not randomly assigned, or self-selection by individuals into exposures. Contact and media/instruction were combined because most contact studies allowed participants to self-select their level of contact and a few studied intact schools that differed in contact, whereas media/instruction studies typically used intact groups or random assignment. Here it is seen that attitude effects were more likely to be found when random assignment (8 of 13 or 62% of effects) and self-selection (17 of 28 or 61%) designs were used (binomial ps < .008 in both cases). Random assignment designs did not lead to a majority of positive outcomes regarding peer relations/behavior (11 of 25 or 44% of effects), nor were positive outcomes significantly greater than would be expected by chance variation (binomial p > .09). Across both attitudinal and relations outcomes, intact groups led to mostly non-significant findings (24 of 38 or 63% of effects), and the number of positive outcomes was not greater than expected by chance (binomial p > .5).

Details on studies using random assignment provide some understanding of why they were more effective than intact groups. Random assignment, especially at the individual level, was used when media and antibias instruction was delivered in a fairly controlled fashion by a researcher. For example, Castelli, De Dea, and Nesdale (2008) showed children a short video of an ingroup man interacting with an outgroup man in a close or distancing fashion; Persson and Musher-Eizenman (2003) and Connolly (Connolly, Fitzpatrick, Gallagher, & Harris, 2006; Connolly & Hosken, 2006) likewise showed videos. Hughes, Bigler, and Levy (2007) and Aboud and Joong (2007) provided antibias instruction on discrimination and how to combat it. So although these studies used strong designs that eliminated potential confounds, exposure was brief and outcomes were measured shortly after. Only Fluent Public Opinion (2008) randomly assigned children to a television program that was delivered in the community with a delayed outcome assessment. In contrast, self-selection into groups of high vs. low exposure was the favored design for contact studies. Exposure in this case was the number of outgroup friends reported by the participant or by classmates; this variable was then correlated with attitudes. Over the course of several months, friends and attitudes converged in most of our studies. This design may lead to more successful outcomes, not because of experimental rigor but because of confounds associated with self-selection, namely that children predisposed to have positive attitudes may make more outgroup friends. Intact groups had two strikes against them: they tended to be groups formed by a non-researcher, such as school personnel, and children's prior proclivities were less likely to determine their exposure to the outgroup.

A clearer picture emerges in Table 5 where effects are counted separately for participants in the majority ethnicity in that setting, e.g., European-Germans, -British, -Americans, and -Canadians. Considering attitudes only, effects for ethnic minority samples were largely non-significant (12 of 14 or 86%, binomial test of positive effects yields p > .86). However, 67% of the attitude effects on majority participants were positive (32 of 48, binomial p < .001), 25% non-significant, and 8% negative. Eight out of 12 effects based on random assignment were positive (67%, binomial p < .004), nine out of 18 based on intact groups were positive (50%, binomial p < .044), and 15 out of 18 based on self-selection were positive (83%, binomial p < .001). Positive effects for peer relations/behaviors remained low at 29% (12 out of 41, with binomial p > .64) when considering ethnic majority participants. Although media/instruction was generally more effective, as first reported, both contact and media/instruction were equally effective with majority participants (data not shown in tables) – 21 of 44 or 48% of contact effects and 23 of 45 or 51% of media/instruction effects were positive for majority ethnic children (both binomial ps < .02; a χ^2 test of the difference yielded $\chi^2 = 0.1$, $df = 1$, p > .74). The highest score of 78% or 18 of 23 effects (binomial p < .001) was found with majority children's attitudes resulting from contact, whereas contact had a mostly non-significant effect on peer relations (14 of 21 effects or 67% were non-significant).

Another interesting variable is the amount of control over delivery of exposure exercised by the researchers. Table 6 shows the positive attitudinal effects for ethnic majority participants using natural exposure (15 out of 18 or 83%, binomial p < .001) and experimental procedures (11 out of 17 or 65%, binomial p < .001), compared to less controlled programs where 7 out of 13 or 54% were positive (binomial p < .035). Considering peer relations/behavior outcomes, 40% of experimental and 27%

Table 3
Effects by outcome (attitude, peer relations) and type of intervention (contact, media/instruction).

	Contact studies $k = 14$				Media and instructional studies $k = 18$			
		Positive	*ns*	Negative		Positive	*ns*	Negative
Attitude $k = 62$ effects	$k = 35$ effects	20 (57%)	14 (40%)	1 (3%)	$k = 27$ effects	14 (52%)	10 (37%)	3 (11%)
Peer reln & beh $k = 59$ effects	$k = 35$ effects	5 (14%)	22 (63%)	8 (23%)	$k = 24$ effects	10 (37%)	14 (52%)	0 (0%)

Table 4
Effects by outcome (attitude, peer reln) and design (random, intact, self-selected into exposure group).

	Attitude effects $k = 62$				Peer reln and behavior effects $k = 59$			
		Positive	*ns*	Negative		Positive	*ns*	Negative
Random assignment to IV groups	$k = 13$ effects	8	3	2	$k = 25$ effects	11	12	2
Intact IV groups not random	$k = 21$ effects	9	10	2	$k = 17$ effects	3	14	0
Self-selected into IV groups	$k = 28$ effects	17	11	0	$k = 17$ effect	1	10	6

of programs yielded positive effects, whereas 22% of natural exposures were beneficial (all binomial *p*s indicating that none of these produced more positive results than expected by chance).

A few examples will demonstrate why natural exposure and experimental procedures were more effective. Natural exposure tended to be in schools where ethnic diversity occurred not by design and children freely chose their own exposure within a diverse student body. Experimental manipulations were usually brief and intense, such as hearing an exchange between an ingroup and an outgroup person (Castelli, De Amicis, & Sherman, 2007) and hearing a bystander intervene to stop nasty name-calling (Aboud & Miller, 2007). In contrast, programs were usually designed and delivered by the community or the school: local television programming in Israel and the West Bank (e.g., Cole et al., 2003), classroom language instruction in French and Spanish (e.g., Aboud & Sankar, 2007; Wright & Tropp, 2005), an after-school integrated program (Aboud, Friedmann, & Smith, in press), multicultural education about different ethnic groups (e.g., Kowalski, 1998; Persson & Musher-Eizenman, 2003), and lessons in social inclusion (e.g., Connolly et al., 2006; Houlette et al., 2004). Scaled-up programs are goals we aim for. Some of their outcomes did show change, but many did not. Researchers often did not have any input on developing the program and there was little formative work to assess on a smaller scale whether outcomes would change. Sometimes the programs were diffuse, mixed in with contradictory instruction during the course of the day about peer and teacher attitudes (Aboud, 2002). Rarely was there a manipulation check on whether the message was received as intended. One study on inclusive identity found that the main identity message was not adopted (Houlette et al., 2004).

Beyond the design and delivery variables, the quality of the program and of the assessment had little impact on the effects (see Tables 7a and 7b). The effects were robust regardless of the measures used, the sample size, adjustment for clustering, and whether there was a manipulation check. Most activities and materials were informed by theory, were age appropriate, and based on prior experimental work, so the main variation in quality of the exposure (program) was who delivered it (the children themselves, the experimenter, or a teacher/parent/media). Many evaluations were conducted with measures of unknown validity, homogeneous samples, infrequent blinding of testers, little control over confounders except through random assignment, and short follow-ups.

Qualities and outcomes of some studies

Several studies from each exposure category are described here briefly to highlight features of some stronger studies and examine why they might have been more or less effective.

Contact studies commonly used the self-selection, natural exposure design, e.g., Aboud, Mendelson, and Purdy (2003) and Feddes, Noack, and Rutland (2009). The Aboud study received a score of 26 for program and 20 for evaluation, while the Feddes study received 24 for program and 21 for evaluation. Both

Table 5
Effects by outcome (attitude, peer relations) and design (random, intact, self-selected into exposure group) for majority/minority ethnic groups.

	Attitude effects $k = 48/14$				Peer reln and behavior effects $k = 41/18$			
		Positive	*ns*	Negative		Positive	*ns*	Negative
Random assignment to IV groups	$k = 12/1$ effects	8/0	2/1	2/0	$k = 18/7$ effects	8/3	10/2	0/2
Intact IV groups not random	$k = 18/3$ effects	9/0	7/3	2/0	$k = 14/3$ effects	3/0	11/3	0/0
Self-selected into IV groups	$k = 18/10$ effects	15/2	3/8	0/0	$k = 9/8$ effect	1/0	4/6	4/2

Table 6
Effects by outcome and exposure control for majority/minority ethnic participants.

	Attitude effects $k = 48/14$				Peer reln and behavior effects $k = 41/18$			
		Positive	*ns*	Negative		Positive	*ns*	Negative
Naturally occurring exposure	$k = 18/10$ effects	15/2	3/8	0/0	$k = 9/8$ effects	2/1	3/5	4/2
Experimental exposure	$k = 17/1$ effects	11/0	4/1	2/0	$k = 10/4$ effects	4/2	6/0	0/2
Program delivered with little control	$k = 13/3$ effects	7/0	5/3	1/0	$k = 22/7$ effects	6/0	16/6	0/0

found positive effects of contact on ethnic majority but not minority children in schools that varied in their level of diversity. The strength of both studies lay in their measurement of contact that went well beyond superficial levels of exposure, and included indirect friends, direct friends, and friendship quality. Attitudes were measured in terms of both positive and negative evaluations. Both studies suffered from the usual limits of contact studies, namely research assistants were not blind, confounders not eliminated, and follow-up was short. Two other studies using intact groups also found positive effects on attitudes (Rutland, Cameron, Bennett, & Ferrell, 2005; Wright & Tropp, 2005). Intact groups, even if not randomly assigned, provide a stronger test of the relation between contact and attitudes than self-selected friendships. Rutland et al. studied preschoolers who attended one of three schools varying in diversity, while Wright and Tropp's students attended one of three classrooms that varied in ethnic diversity and language of instruction. Their program and evaluation quality scores were 25 and 17 for Rutland and 25 and 19 for Wright. The programs were strong because they were informed by theory, yet conducted in a natural setting that could be scaled up. Evaluation marks were lower for these two studies because data collectors were not blind to condition and measures with unknown validity were used. Once again, Wright and Tropp found positive effects on ethnic majority but not minority children (see Tropp & Prenovost, 2008), and Rutland was able to assess only the former. Despite the difference in design, the findings of these four studies are consistent with each other but different from the next study to be described that used a more controlled random assignment design.

Only one contact study used a design with (approximate) random assignment. Leman and Lam (2008) was scored 23 for program and 20 for evaluation quality. Unlike the previously described contact studies, this one used an experimental analogue of contact by inviting pairs of children to attend a session in a private room. The lower program score was therefore due to its short duration in a setting that was not quite child-friendly. Children were paired with either a same-ethnic or cross-ethnic peer to discuss and then select a playmate from among three ethnic groups. The exposure was therefore controlled by the researchers and the outcome was peer relations (playmate selection) and affiliative behavior toward the partner. Ethnic majority children did not show more outgroup peer preference or affiliative behavior when paired with a cross-ethnic partner. The South Asian and British Caribbean children showed more outgroup peer preference but less affiliative behavior when paired with a cross-ethnic compared to a same-ethnic partner. Their greater number of outgroup preferences was matched by an equally greater number of no playmate selections, i.e., they could not come to an agreement with their British White partner. Thus ethnic minority children either complied with the majority child's request for an ingroup playmate or resisted the request, resulting in a stand-off. These five

Table 7a
Effects by outcome by quality of program (i.e. delivery of the exposure variable) nine qualities rated 1–3.

	Attitude effects $k = 62$				Peer reln and behavior effects $k = 59$			
		Positive	*ns*	Negative		Positive	*ns*	Negative
Quality low 17–22	$k = 23$ effects	12	8	3	$k = 10$ effects	5	5	0
Quality medium 23–24	$k = 22$ effects	14	7	1	$k = 36$ effects	5	23	8
Quality high 25+	$k = 17$ effects	8	9	0	$k = 13$ effects	5	8	0

Table 7b
Effects by outcome by quality of assessment (i.e. evaluation of effectiveness) nine qualities rated 1–3.

	Attitude effects $k = 62$				Peer reln and behavior effects $k = 59$			
		Positive	*ns*	Negative		Positive	*ns*	Negative
Quality low < 18	$k = 23$ effects	13	8	2	$k = 26$ effects	2	18	6
Quality medium 18–19	$k = 17$ effects	9	6	2	$k = 14$ effects	6	8	0
Quality high 20+	$k = 22$ effects	12	10	0	$k = 19$ effects	7	10	2

contact studies had relatively high quality scores and demonstrated the typical findings that self-selection and natural exposure were more effective than random assignment, and that attitudes were more likely to change than peer relations and behavior.

Two studies using media (viz. Cameron, Rutland, Brown, & Douch, 2006; Fluent, 2008) and two studies conveying some form of antibias instruction (viz., Houlette et al., 2004; Hughes et al., 2007) used random assignment. One of each used an experimenter-controlled intervention: Cameron and colleagues modified several story books about cross-ethnic friends and the experimenter read these to small groups of children, while Hughes developed biography materials and instructed children about overcoming racial discrimination. The other two were delivered in less controlled settings by parents and teachers: Fluent evaluated the impact of viewing at home a minimum 12 episodes of Sesame Street in Serbia and Kosovo, and Houlette's students participated in a program delivered by school personnel, that was first developed by an organization and then enhanced by researchers. Quality totals here ranged from 20 to 25 for the program and 19 to 23 for the evaluation. The two media studies reported positive outcomes: children who heard the stories or watched the televised episodes were more positive to outgroup members than a control group with no input. Cameron's study is limited by having unscripted discussions with children after story-reading, whereas Fluent's study said little about blinding or validity of the measures. However, both were informed by theory, namely a media-mediated form of contact theory that is particularly appropriate for children who have little opportunity for direct contact. Both could easily be implemented on a large scale.

The Hughes et al. study showed a more positive evaluation of outgroup members among majority (European American) children after hearing about famous African American figures and their experiences with unjustified discrimination. The intervention itself was short, follow-up immediate, and research assistants were not blind to condition. However, it experiments with a new perspective on prejudice reduction, namely exposing children to egregious forms of bias that arouse anger and an antibias response (Finlay & Stephan, 2000). These concepts were raised by Derman-Sparks and Philips (1997) in their writings on early childhood education and have been applied in anti-bullying research (e.g., Aboud & Miller, 2007). The study by Houlette showed minimal effects, even for the enhanced program, and used unconventional measures of peer relations. The sample size was large, though classes rather than individuals were the unit of analysis, to accommodate clustering. This last study produced findings that were common among instructional programs delivered in a school setting and developed initially by educators (see also Connolly et al., 2006; Kowalski, 1998; Perkins & Mebert, 2005; Wham, Barnhart, & Cook, 1996). They are often well-intentioned but have not undergone the small-scale experimental work with children to determine if they produce positive outcomes under controlled conditions. We found that in many cases the proposed mediating variable, such as empathy or inclusive identity, had not changed after the intervention.

Appendix A

Summary of findings and quality assessment (Under Effects, X = Attitude outcome, Y = Peer relation/behavior outcome).

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	Effects 0	Effects –	Quality of Prog	Quality of Assm't
Contact (k = 14)											
Aboud et al. (2003)	Euro- *n* = 164, Caribbean *n* = 76 Canadian	6–12	Naturally occurring	Diverse student body (high diversity)	Self-selected, Cross-sec	Evaluations of *Caribbean by Euro*				S8 M1 W0 26	S4 M3 W2 20
						excluded	X				
						companion	X				
						friend		X			
						quality of friend	X				
						Evaluations of *Euro by Caribbean:*					
						excluded		X			
						companion		X			
						friend		X			
						Quality of friend		X			
Aboud & Sankar (2007)	Francophone *n* = 44 Anglophone *n* = 41 Canadian	8–11	Program variation diversity between Schools	Diverse student body (high diversity), Language instruction	Intact groups *n* = 44 vs. 41, Cross-sec	Peer relations with *Anglophone:*				S7 M1 W1 24	S2 M4 W3 17
						Companion	Y				
						Friend		Y			
						with *Francophone:*					
						Companion		Y			
						Friend		Y			
Aboud et al (in press)	Francophone *n* = 74 Anglophone *n* = 74 Canadian	7–10	Program variation in diversity	Separate or Integrated, Language instruction	Intact groups *n* = 93 vs. 55, Cross-sec	Direct friend		Y		S7 M1 W1 24	S2 M4 W3 17
						indirect friend		Y			
						Evaluation			X		

(*continued on next page*)

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	Effects 0	Effects –	Quality of Prog	Quality of Assm't
Castelli et al. (2007)	Euro- Italian *n* = 36, 120, 42, 63	3–5	Naturally occurring	Contact (self-rated)	Self-selected, Cross-sec	Evaluation of *African-Italian*				S3	S3
						Study 1	X			M4	M3
						Study 2	X			W2	W3
						Study 3		X		19	18
						Study 4	X				
Farhan (2008)	Euro- Germans *n* = 127	8–12	Naturally occurring	Friendship (self-rated)	Self-selected, Longitud	Evaluation of *Turkish-German*	X			S7	S3
						Immigrants	X			M1	M4
										W1	W2
										24	19
Feddes et al. (2009)	Euro- *n* = 104, Turkish *n* = 98 Germans	7–11	Naturally occurring	Friendship (self-rated) in a diverse student body (low diversity)	Self-selected,	Evaluations of *Turkish-German*				S7	S4
						With direct fr:				M1	M4
					Cross-Sec, Longitud,	-T1 cross-sect		X		W1	W1
						-T2 cross-sect	X			24	21
						-T2 longitudinal	X				
						With indirect fr:					
						-T1 cross-sect	X				
						-T2 cross-sect	X				
						-T2 longitudinal	X				
						Evaluations of *Euro-German*					
						With direct fr:					
						-T1 cross-sect		X			
						-T2 cross-sect		X			
						-T2 longitudinal		X			
						With indirect fr:					
						-T1 cross-sect	X				

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	0	−	Quality of Prog	Quality of Assm't
						-T2 cross-sect	X				
						-T2 longitudinal		X			
Finkelstein & Haskins (1983)	Euro- *n* = 25, African n = 38 Americans	5–6	Naturally occurring	Diverse student body (high diversity)	Self-select, Longitud	Peer relations with *African-Amer*				S6	S1
						Recess friend	Y			M3	M4
						-talk		Y		W0	W4
						-negative beh			Y	24	15
						Work friend		Y			
						-talk			Y		
						-negative beh			Y		
						with *Euro-Amer*					
						Recess friend			Y		
						-talk		Y			
						-negative talk		Y			
						Work friend		Y			
						-talk		Y			
						-negative beh		Y			
Graham (1998)	Euro- *n* = 77, African *n* = 68 American	6–12	Naturally occurring	Diverse student body (high diversity)	Self-selected, Longitud	Friendship with *African Americans*				S6	S1
										M3	M5
										W0	W3
						-boys		Y		24	16
						-girls			Y		
						with *Euro-Amer*					
						-boys		Y			
						-girls			Y		

(*continued on next page*)

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	Effects 0	Effects –	Quality of Prog	Quality of Assm't
Howes & Wu (1990)	Euro- *n* = 127, African *n* = 36, Hispanic *n* = 33, Asian *n* = 14 American	5–9	Naturally occurring	Diverse student body (low diversity)	Self-selected, Cross-sec	Peer interaction		Y		S5 M4 W0 23	S4 M3 W2 20
Leman & Lam (2008)	White *n* = 226, Caribbean *n* = 99, South Asian *n* = 103 British	7–8	Expt'l manipulate	Dyadic inter-ethnic conversation	Random individual, post, Ingroup White *n* = 55, ingroup minority *n* = 68, mixed dyad *n* = 68	Affiliation with Friendship with *South Asian* by Euro-British Affiliation with Friendship with *African-Caribbean* By Euro-British Affiliation with Friendship with *Euro-British* by South Asian Affiliation with Friendship with *Euro-British* by African Caribbean	 Y Y	Y Y Y 	 Y Y	S6 M2 W1 23	S5 M1 W3 20
McGlothlin & Killen (2006)	Euro- *n* = 168 Americans	6–11	Program	Diverse student body (high diversity)	Intact groups, Cross-sec, *n* = 74 vs. 94	Evaluation of *African Americans*	X			S7 M1 W1 24	S1 M5 W3 16

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	0	−	Quality of Prog	Quality of Assm't
Rutland et al. (2005)	Euro-British $n = 136$	3–4	Program	Diverse student body (high diversity)	Intact groups, Cross-sec, $n = 63$ vs. 45 vs. 28	Evaluation of *African-Caribbean*	X			S7 M2 W0 25	S3 M2 W4 17
Turner et al. (2007)	Euro-British $n = 60$	8–12	Naturally occurring	Friendship (self-rated in a diverse neighborhood) (low diversity)	Self-selected, Cross-sec	Explicit evaluation of *South Asian*	X			S4 M1 W4 18	S3 M2 W4 17
						Implicit evaluation of *South Asian*	X				
Wright & Tropp (2005)	Euro-Anglo $n = 356$, Hispanophone $n = 484$ American	5–9	School program	Diverse student body (high diversity) plus Language instruction	Intact groups of Euro-Anglo, Cross-sec,$n = 163$ vs. 80 vs. 113	Evaluations of *Latino-American* Due to ethnic integration	X			S7 M2 W0 25	S3 M4 W2 19
						Due to language of instruction		X			
						Friend preference Due to ethnic integration		Y			
						Evaluations of *Euro-American* Due to language of instruction	Y				
						Evaluations of *Euro-American* Due to ethnic integration		X			

(*continued on next page*)

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	0	−	Quality of Prog	Quality of Assm't
						Due to language of instruction		X			
						Friend preference					
						Due to ethnic integration		Y			
						Due to language of instruction		Y			
Media-mediated (k = 9)											
Aboud (2002)	Euro-Canadian *n* = 59	5–8	Expt'l manipulate	4 storybooks about intergroup friends read by a Euro- or Caribbean Canadian adult	Intact group, pre-post *n* = 34 vs. 25	Evaluation of *Caribbean Can*	X			S6 M3 W0 24	S2 M2 W5 15
						Inferred Evaluation of *CaribbeanCan*					
						By Euro- K		X			
						Grade1	X				
						By Caribbean		X			
Cameron et al. (2006)	Euro-British *n* = 253	5–12	Expt'l manipulate	3 storybooks about intergroup friends	Random cluster, post *n* = 67,66, 66 vs. 54	Evaluation of refugees	X			S5 M4 W0 23	S5 M3 W1 22
						Social distance		Y			
Cameron et al. (2007)	Euro-British *n* = 198	6–11	Expt'l manipulate	3 storybooks about intergroup friends	Random cluster, post *n* = 50,50, 50 vs. 48	Evaluation of *refugees*	X			S5 M4 W0 23	S5 M3 W1 22
						Social distance		Y			
Castelli et al. (2008)	Euro-Italian	3.5–6.5	Expt'l	Euro-Italian &	Random	Ch's evaluation	X			S5	S3

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	0	−	Quality of Prog	Quality of Assm't
	n = 79		manipulate	African-Italian in face-to-face video interaction vary in non-verbal positivity	individual post, *n* = 40 vs. 39	Ch Generalized liking of *African Italians*	X			M2 W2 21	M2 W4 17
						Inferred helping of *African Italian*		X			
						Perc'd positive relationship	X				
						Generalized inferred helping of *African-Italian*	X				
Cole et al. (2003)	Jewish Israeli *n* = 110, Palestinian Israeli *n* = 63, Palestinian *n* = 99	3–6	Video program	Television program in local language showing Jewish and Palestinian life plus educational messages	Intact groups, pre-post	Evaluations of *Arabs*	X			S7 M2 W0 25	S2 M4 W3 17
						Problem solving by Jewish Israeli Of *Jews* by Palestinian Israeli	X	Y			
						Problem solving Of Jews by Palestinians		Y X			
						Problem solving		Y			
Fluent (2008)	Serbian *n* = 264, Albanian *n* = 272 Kosovans	5–6	Video program	Television program in local language showing live action,	Random individual, post Serbian *n* = 119 vs.	Social distance to *Chinese*	Y(S)	Y(A)		S7 M2 W0 25	S6 M2 W1 23
						Social distance to *Roma*	Y				
						Social inclusion					

(*continued on next page*)

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	Effects 0	Effects –	Quality of Prog	Quality of Assm't
				animation, ethnic messages	145, Albanian $n = 134$ vs. 138	of other language gp	Y(S)	Y(A)			
Kowalski (1998)	Euro- $n = 30$ American	3–5	School program	Video, stories, letters, drawings from Japanese preschool	Intact groups, pre-post $n = 18$ vs. 12	Forced choice Free choice Friend preference		X Y	X	S5 M2 W2 21	S1 M5 W3 16
Persson & Musher-Eizenman (2003)	Euro- $n = 60$ American	3–6	Expt'l manipulate	Video stories of intergroup playmates and talk of respect	Random individual, pre-post, $n = 45$ vs. 15	Evaluations Liking		X Y		S5 M3 W1 22	S4 M1 W4 18
Wham et al. (1996)	Euro- $n = 128$ American	6–10	School program	Intra-ethnic stories of children	Intact groups, pre-post $n = 69$ vs. 59	Evaluations by K by Grade 2	 X	X		S5 M3 W0 21	S2 M4 W3 17
Antibias & Multiculture (k = 9)											
Aboud & Doyle (1996)	Euro- $n = 88$ Canadian	8–11	Expt'l manipulate	Dyadic intra-ethnic conversation about ethnic evaluations	Intact groups, pre-post	Evaluation of *Caribbean Canadian*	X			S4 M4 W1 21	S3 M3 W3 18
Aboud & Joong (2007)	Diverse ethnicity $n = 73$	8–12	Expt'l manipulate	Audio antibias models; Diverse student body (high diversity)	Random individual, pre-post, $n = 36$ vs. 37	Assertive antibias talk to name-caller Moral rationale for not name calling	Y Y			S5 M3 W1 22	S3 M4 W2 19

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	0	−	Quality of Prog	Quality of Assm't
Aboud & Miller 2007	Diverse ethnicity *n* = 96	8–12	Expt'l manipulate	Audio antibias models; Diverse student body (high diversity)	Random individual, pre-post, *n* = 48 vs. 48	Assertive antibias talk to name-caller Moral rationale for not name calling	Y Y			S5 M3 W1 22	S2 M5 W2 18
Connolly et al. (2006)	Catholic & Protestant North Irish *n* = 165	3–4	School program	Videos & classroom materials of inclusive models	Random clusters, pre-post *n* = 95 vs. 70	Like to play with -all outgroups -religious outgroup	Y	 Y		S5 M3 W1 22	S4 M3 W2 20
Connolly & Hosken (2006)	Catholic & Protestant North Irish *n* = 174	6–7	School program	Theater & classroom materials of inclusive models	Random clusters, pre-post *n* = 87 vs. 87	Like as friend Happy to be friends		Y Y		S5 M3 W1 22	S4 M2 W3 19
Houlette et al. (2004)	Diverse American ethnicity *n* = 700	6–7	School program	Curriculum & materials on inclusion; Diverse student body (high diversity)	Intact groups, pre-post *n* = 35 vs. 17 vs. 9 using class as the unit	Happy to be friends Friend preference Sharing resources	 Y	Y Y		S6 M3 W0 24	S3 M4 W2 19

(*continued on next page*)

Appendix A (*continued*)

Author (Date)	Ethnicity of Participants	Age of Ss (yrs)	Context of Exposure	Indep variable: Form of Exposure to outgroup	Design	Outcome measure	Effects +	Effects 0	Effects –	Quality of Prog	Quality of Assm't
Hughes et al. (2007)	Euro- and African-American *n* = 48	6–11	Expt'l manipulate	Materials & instruction on biography & discrimination	Random individual pre-post *n* = 24 vs. 24	Evaluation of *African American* Evaluation of *Euro-American*	X	X		S3 M5 W1 20	S3 M4 W2 19
Perkins & Mebert (2005)	Euro- *n* = 79 American	3–6	School program	Materials & instruction on cultures	Intact groups, post-only *n* = 43 vs. 36	Evaluations of *Japanese* Evaluations of *African American*		X X		S4 M0 W5 17	S3 M2 W4 17
Verkuyten & DeWolf (2007)	Euro-Dutch *n* = 337	6–10	Expt'l manipulate	Antibias stories	Random individual, post-only *n* = 170 vs. 167	Evaluation of *Moorlanders* Evaluation of *Chinese* Evaluation of *Americans*	X		X X	S3 M4 W2 19	S4 M2 W3 19

Notes:
Ethnic terms *n*'s refer to sample sizes of participants from each ethnic group if analyzed separately
Context of exposure:
Naturally occurring among individuals if school if as a whole is integrated (diverse)
Experimental manipulation
Programs: planned or natural variation in exposure through a specific program offered
Exposure to outgroup is high diversity if 30% or more in the school.
Design *n*'s refer to sample sizes of intervention and control groups.
Quality of Program and Assessment: S = strong, M = medium, W = weak.

Discussion

Our review of 32 studies on interventions delivered to young children found that somewhat fewer than half of the effects were positive. Very few had negative effects. Attitudes were more strongly and positively affected than peer relations. In particular, in both self-selection and random assignment designs, approximately 60% of the attitude effects were positively associated with the exposure. Self-selection designs were used mostly to study the effects of naturally-occurring contact, whereas random assignment was almost exclusively used by media or instructional interventions. The other common design, namely use of non-randomly assigned intact groups who received different levels of exposure, was seen across different types of intervention and yielded only 40% positive effects. Consequently, effects varied by outcome and design. This distinction will receive fuller interpretation shortly.

Outcomes were also more positive for majority ethnic children than for minority ethnic ones when attitudes, but not peer relations, were considered. Overall, 67% of the majority child outcomes were positive, and this percentage was higher for self-selection designs. This finding was not entirely unexpected, given previous reviews (Tropp & Pettigrew, 2005). In fact, many researchers take the view that the problems of prejudice and exclusion are more prevalent and serious in majority ethnicities. They may develop their intervention from research on the correlates of prejudice and discrimination as evidenced by majority children, which may not apply to minority children. While 15 of the 32 studies were confined to majority ethnic children only, and 17 included both majority and minority samples, a larger proportion of the analyzed effects were based on the majority ethnic data – 89 effects for majority vs. 32 for minority. Positive effects on one measured outcome may have led to further analyses, whereas non-significant effects for minority samples may have stopped further analyses.

For minority samples, 72% of effects were not significant. This held for both contact and media/instructional types of intervention. Some researchers have explained the lack of effects on minority children on the basis that they are starting with a positive outgroup orientation in many cases and so have little room to improve (Aboud, 2008). An alternative explanation is that the first priority of all children is to develop a positive attitude and orientation to ingroup members, and only then will they attend to outgroup information (e.g., Aboud, 2003; Patterson & Bigler, 2006). So we might expect to find more effects of exposure to ingroup information among very young children of all ethnicities and of minority children in particular (e.g., Cole et al., 2003).

Both contact and media/instruction forms of intervention showed some success – media slightly more successful at 47% compared to contact with 36% success. The highest percentage of positive effects was seen in the attitudes of majority children during contact where 78% were positive. While reviews across the age span have shown positive effects of contact (Pettigrew & Tropp, 2006), no prior systematic reviews have integrated the effects of media or antibias instruction. These two latter interventions are more recent than contact and so require fine-tuning and a focus on theory such as education–entertainment (Slater & Rouner, 2002) and media-mediated indirect contact (Mutz & Goldman, 2010). Media is a particularly convenient way of providing children with an indirect or vicarious form of contact, especially children with little or no opportunity for direct contact. Media may take the form of books or video and may be mass media or small media. The media studies presented children with scenes and stories of intergroup contact among peers; this is a media form of indirect contact, in which the child identifies with an ingroup peer who has an outgroup friend. In contrast, multicultural education normally focuses on the outgroup culture, and the few multicultural studies in our review did not fare well. Antibias instruction, in contrast, exposes the child to information on incidents of exclusion, and provides suggestions on emotional and behavioral responses to the incident. It aims to directly tackle prejudice and discrimination by coaching children on how to respond when they encounter it in their own or other groups.

Why should attitudes be more positively affected than peer relations? There is a long history of attitude measurement, and the scales used here were commonly based on positive and negative evaluations that have been found reliable and valid indicators of attitudes. In contrast, the measures of peer relations ranged from friend nominations, and liking ratings to social distance (e.g., intention to interact) and agreeing on a playmate. Still, there was no consistent difference among these, such as self-report of friendship or actual interaction, in terms of degree of change. Because of the many

potential determinants of peer relations, including the availability and willingness of another child to be a friend, effects may be harder to find in the short term. Finally, we note that peer relations were often used in the research as an indicator of contact to which attitudes were correlated. As these measures are becoming more sophisticated, researchers should think of including them along with attitudes, as outcome variables. It is interesting to note that in developmental research, attitudes tend to become more expansive with age whereas peer relations become more constricted (e.g., Aboud, 2008); interventions at this young age tend to have the same effects.

Self-selection was generally a more beneficial way to access exposure than being part of an intact group that received high or low exposure. Specifically, 83% of the self-selection designs with majority samples yielded positive attitudinal outcomes. Self-selection was a common mechanism of exposure in the contact studies. The limitations of self-selection are well known: children with positive predispositions (if not a positive attitude at baseline) may seek close contact with outgroup members if the school provides diversity. Attitudes therefore may nudge contact rather than contact leading to positive attitudes. Furthermore, participants may want to be consistent in their responses, thus reporting attitudes to be consistent with friendship. Some studies were quite transparent in their focus on ethnicity by asking for names of friends from the outgroup; this might heighten the desire to be consistent (Davies et al., 2011). The fact that intact groups showed much less difference due to contact than self-selected groups warns us to show caution in our conclusions. We cannot conclude with confidence that contact leads to respect and inclusion among young children until findings from randomized or well controlled non-randomized designs become available.

Initially, we referred to the perception of a gap between research and practice (Paluck & Green, 2009). Experimental interventions with high researcher control indeed yielded positive results for attitude change among ethnic majority children (65%). A similar number of full-fledged programs implemented by school or community personnel also produced positive outcomes but slightly fewer (54%). Experimental, controlled interventions may be more theory-based and faithful to the intended intervention than programs, but the two did not differ in the overall quality of their intervention or evaluation. So it is feasible to implement high quality programs and conduct high quality evaluations of them. The message for media and antibias instruction interventions, therefore, is to move carefully from the experimental analogue to the manualized program. The message for school program designers with little knowledge of the background literature or theory is to make use of program quality criteria to guide their planning of the intervention. They should conduct small-scale formative and pilot research to assess children's acceptance of the materials and activities. Finally, institutions and organizations planning programs could collaborate with researchers to design and conduct an independent evaluation that would help recommend improvements to the program.

Limitations of the review

A limitation of the review concerns the analysis of studies. Because a meta-analysis was not appropriate (see our earlier remarks in this respect), we estimated the strength of findings by comparing the number of effects that were positive, non-significant, or negative. Although large sample sizes may provide greater confidence in certain findings, they also make it easier to find significance, and we therefore did not weight findings by sample size. On inspection, we found that sample size and overall quality of the study did not appear to be related to the presence of positive effects. The only quality that was influential concerned the assignment of participants to condition, where random assignment and self-selection led to a higher rate of positive effects.

A second limitation was our inability to focus solely on children 8 years and under. Most studies did not disaggregate by this age category, so effects were often reported for a broader age range that might include children older than 8 years. Other variables known to influence outcomes, such as the ethnicity of participants, were sometimes considered when tallying effects. However, many studies did not disaggregate by ethnicity or gender either because the attribute was found to be non-significant in preliminary analyses or participants were not asked to self-identify.

A third limitation was the inability to study mediators of exposure effects, given the small number and heterogeneity of studies. Theories have identified cognitive and behavioral skills that are acquired as a result of contact, media and classroom instruction. Recent research now includes an assessment of

anxiety, perceived norms, and attention to individual differences, but rarely with young children. Ways to measure these mediating variables in a child-friendly manner is ongoing (Patterson & Bigler, 2006).

Implications for theories of change

Our overall impression was that interventions explicitly informed by a theory were stronger than ones informed by good intentions, or only implicitly by theory. Contact theory was most often identified as the basis for an intervention, and recently the extended contact theory was tested where an outgroup member is one-step removed (e.g., Cameron et al., 2006; Turner, Hewstone, & Voci, 2007). A few studies sought out intact groups that varied in ethnic and language contact (e.g., Rutland et al., 2005; Wright & Tropp, 2005). Our conclusions about support for contact theory echo those of others (e.g., Pettigrew & Tropp, 2006; Raabe & Beelmann, 2011). However, we also identified the problem of drawing conclusions from self-selected exposure to outgroups. When students attended a mixed-ethnicity school, their level of contact was usually self-reported and this in turn was correlated with attitudes. The effects of these studies were more likely to be positive than ones where children were assigned to a high or low contact school. Consequently, a stronger test of contact theory requires assignment to different levels of contact, not based on the child's predisposition (e.g., Cameron et al., 2006; Leman & Lam, 2008; Rutland et al., 2005; Wright & Tropp, 2005).

Media studies, presenting children with audio-visual or print material, sometimes used extended contact theory (e.g., Cameron et al., 2006) or learning theory (e.g., Castelli et al., 2008) to guide their intervention. Some of these interventions were loosely connected to antibias theory in that they exposed children to exclusion or name-calling; they then relied on observational learning and a social-construction approach to encourage children to construct a response they would feel comfortable expressing in such situations. This research may need to broaden its use of theory to include communication theory in its current form; the Elaboration Likelihood Model and its application to entertainment–education identify aspects of the receiver that need to be considered when designing the message.

Antibias theory (e.g., Bigler & Liben, 2006) explicitly considers developmental capabilities of the receiver that facilitate or impede accurate processing of the message. It notes why simply presenting positive information of an outgroup is insufficient. Messages need to be tailored to the cognitive and emotional maturity of the children who in most cases already have well-formed opinions of themselves and others. Although many of these propositions have been known for decades (e.g., Aboud, 2008; Katz & Zalk, 1978), they have not been incorporated into interventions.

Implications for future research

The preceding paragraphs underscore our need for more research on media and instructional interventions. Both provide easy formats for manipulating theoretically interesting variables in a systematic design. These include a comparison of antibias vs. one-sided positive messages about an outgroup, extended contact vs. outgroup-only messages, ingroup vs. outgroup communicators, and prior training to overcome egocentric or dichotomous thinking. Although we aim for programs that can be scaled up throughout a school or community, many programs are currently being implemented without prior scrutiny to ensure that they are theory- and evidence-based. So we advocate a rigorous experimental analogue of media/instruction before implementing on a larger scale.

The quality of evaluations would be improved greatly with the use of random assignment of clusters or individuals in order to eliminate confounds associated with self-selection. If random assignment is not feasible, as with school programs, a pretest assessment of groups should be conducted to demonstrate that no initial differences exist or that differences are controlled statistically. Blinding of research assistants should also be ensured in order to eliminate bias. Conventional measures of attitudes and peer relations are now readily available and can easily be modified for age and intergroup context. Finally a manipulation check would allow for confidence that the intended message was received. We should not be surprised that children do not always receive the message about positive outgroup evaluations when a story is read by an ingroup adult (Aboud, 2002) or about an inclusive identity when extracurricular activities emphasize an exclusive identity (Houlette et al., 2004).

We found our search for relevant articles difficult because authors did not use expected keywords such as *prejudice reduction* or *cross-ethnic peer relations* or *program evaluation*. Coming from different disciplines means that we need to coordinate our research with common keywords. Also, our assessment of quality was sometimes difficult because articles were not always explicit about the theory and evidence base of the intervention, the control group's exposure, the reliability and validity of the outcome measures, and the characteristics of the sample.

In conclusion, this systematic review of 32 studies revealed valuable patterns in the findings of contact, media and instructional interventions delivered to young children. Media and contact were the most effective; instructional forms of antibias showed promise. Both media and instructional formats need more research, in particular using random assignment designs. Attitudinal outcomes showed greater change than peer relations, and majority ethnic children were more influenced by these interventions than minority ethnic children. These findings draw attention to promising avenues for future research by highlighting the characteristics of interventions that should be replicated and extended. They also underscore our need to improve the quality of research designs and systematically include theories of change that are relevant to early childhood.

Acknowledgments

We thank additional members of the Una Global Evaluation group, Charlotte Cole, Buhle Zuma, and Luigi Castelli, for their contribution. We are grateful to Paul Connolly and the Una Global Organization for providing funds to support this review, through the Bernard van Leer Foundation, and Atlantic Philanthropies (www.unaglobal.org).

References

*Denotes the study is included in the systematic review

*Aboud, F.E. (2002). Antibias messages and prejudice reduction in young children. Paper presented at the Society for the Psychological Study of Social Issues, Toronto, June.

Aboud, F. E. (2003). The formation of ingroup favoritism and outgroup prejudice in young children: Are they distinct attitudes? *Developmental Psychology, 39*, 48–60.

Aboud, F. E. (2005). The development of prejudice in childhood and adolescence. In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport* (pp. 310–326). Malden, MA: Blackwell.

Aboud, F. E. (2008). A social-cognitive developmental theory of prejudice. In S. M. Quintana & C. McKown (Eds.). *The handbook of race, racism, and the developing child* (pp. 55–71). Hoboken, NJ: Wiley.

*Aboud, F. E., & Doyle, A. B. (1996). Does talk of race foster prejudice or tolerance in children? *Canadian Journal of Behavioural Science, 28*, 161–170.

*Aboud, F. E., Friedmann, J., & Smith, S. (in press). Direct and indirect friends in cross-ethnic peer relations. *Social Development.*

*Aboud, F. E., & Joong, A. (2007). Intergroup name-calling and conditions for creating assertive bystanders. In S. R. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 249–260). Oxford: Oxford University Press.

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. 269–293). Mahwah, NJ: Erlbaum.

*Aboud, F. E., Mendelson, M. J., & Purdy, K. T. (2003). Cross-race peer relations and friendship quality. *International Journal of Behavioral Development, 27*(2), 165–173.

*Aboud, F. E., & Miller, L. (2007). Promoting peer intervention in name-calling. *South African Journal of Psychology, 37*, 803–819.

*Aboud, F. E., & Sankar, J. (2007). Friendship and identity in a language-integrated school. *International Journal of Behavioral Development, 31*, 445–453.

Abrams, D., Rutland, A., & Cameron, L. (2003). The development of subjective group dynamics: Children's judgments of normative and deviant in-group and out-group individuals. *Child Development, 74*, 1840–1856.

Allport, G. W. (1954/1979). *The Nature of Prejudice*. Cambridge, MA: Perseus Books.

Banks, J. A. (Ed.). (2009). *The Routledge international companion to multicultural education*. New York: Routledge.

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in child development and behavior (Vol. 34)* (pp. 39–89). San Diego: Elsevier.

Brown, R. (1995). *Prejudice. Its social psychology*. Cambridge, MA: Blackwell.

Cameron, L., & Turner, R. N. (2010). The application of diversity-based interventions to policy and practice. In R. J. Crisp (Ed.), *The psychology of social and cultural diversity*. Blackwell. doi:10.1002/9781444325447.

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

*Cameron, L., Rutland, B., & 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*(5), 454–466.

*Castelli, L., De Amicis, L., & Sherman, S. J. (2007). The loyal member effect: On the preference for ingroup members who engage in exclusive relations with the ingroup. *Developmental Psychology, 43*, 1347–1359.

*Castelli, L., De Dea, C., & Nesdale, D. (2008). Learning social attitudes: Children's sensitivity to the nonverbal behaviors of adult models during interracial interactions. *Personality & Social Psychology Bulletin, 34*, 1504–1513.
*Cole, C. F., Arafat, C., Tidhar, C., Tafesh, W. Z., Fox, N. A., Killen, M., et al (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, West Bank, and Gaza. *International Journal of Behavioral Development., 27*, 409–432.
*Connolly, P., Fitzpatrick, S., Gallagher, T., & Harris, P. (2006). Addressing diversity and inclusion in the early years in conflict-affected societies: a case study of the Media Initiative for Children —Northern Ireland. *International Journal of Early Years Education, 14*, 263–278.
*Connolly, P., & Hosken, K. (2006). The general and specific effects of educational programmes aimed at promoting awareness of and respect for diversity among young children. *International Journal of Early Years Education, 14*, 107–126.
Davidson, K. W., Goldstein, M., Kaplan, R. M., Kaufmann, P. G., Knatterund, G. L., Orleans, C. T., et al (2003). Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine, 26*, 161–171.
Davies, K., Tropp, L. R., Aron, A., Pettigrew, T. F., & Wright, S. C. (2011). Cross-group friendships and intergroup attitudes: A meta-analytic review. *Personality & Social Psychology Review, 15*, 332–351.
Deeks, J. J., Dines, J., Amico, R. D., Sowden, A. J., Sakarovitch, C., Song, F., et al (2003). Evaluating non-randomized intervention studies. *Health Technology Assessment, 7*, 1–192.
Derman-Sparks, L., & Philips, C. B. (1997). *Teaching/learning anti-racism: A developmental approach*. New York: Teachers College, Columbia University.
Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health, 52*, 377–384.
Ellison, C. G., & Powers, D. A. (1994). The contact hypothesis and racial attitudes among Black Americans. *Social Science Quarterly, 75*, 385–400.
*Farhan, T. (2008). *Prejudice in school children: Causal analyses of possible factors of influence based on a longitudinal study*. Germany: Philipps University Marburg.
*Feddes, A. R., Noack, P., & Rutland, A. (2009). Direct and extended friendship effects on minority and majority children's interethnic attitudes: A longitudinal study. *Child Development, 80*(2), 377–390.
*Finkelstein, N. W., & Haskins, R. (1983). Kindergarten children prefer same-color peers. *Child Development, 54*, 502–508.
Finlay, K. A., & Stephan, W. G. (2000). Improving intergroup relations: The effects of empathy on racial attitudes. *Journal of Applied Psychology, 30*, 1720–1737.
*Fluent Public Opinion (2008). Assessment of educational impact of Rruga Sesam and Ulica Sezam: Report of findings prepared for Sesame Workshop, New York.
*Graham, J. A. (1998). A longitudinal investigation of race and sex as factors in children's classroom friendship choices. *Child Study Journal, 28*, 245–266.
*Houlette, M. A., Gaertner, S. L., Johnson, K. M., Banker, B. S., Riek, B. M., & Dovidio, J. F. (2004). Developing a more inclusive social identity: An elementary school intervention. *Journal of Social Issues, 60*, 35–55.
*Howes, C., & Wu, F. (1990). Peer interactions and friendships in an ethnically diverse school setting. *Child Development, 61*, 537–541.
*Hughes, J. M., Bigler, R. S., & Levy, S. R. (2007). Consequences of learning about historical racism among European American and African American children. *Child Development, 78*, 1689–1705.
Katz, P. A., & Zalk, S. R. (1978). Modification of children's racial attitudes. *Developmental Psychology, 14*, 447–461.
Kenworthy, J. B., Turner, R. N., Hewstone, M., & Voci, A. (2005). Intergroup contact: When does it work, and why? In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), *On the nature of prejudice. Fifty years after Allport* (pp. 278–292). Malden, MA: Blackwell.
*Kowalski, K. (1998). The impact of vicarious exposure to diversity on preschoolers' ethnic/racial attitudes. *Early Child Development & Care, 146*, 41–51.
*Leman, P. J., & Lam, V. L. (2008). The influence of race and gender on children's conversations and playmate choices. *Child Development, 79*, 1329–1343.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
*McGlothlin, H., & Killen, M. (2006). Intergroup attitudes of European American children attending ethnically homogeneous schools. *Child Development, 77*, 1375–1386.
Mutz, D. C., & Goldman, S. K. (2010). Effects of mass media on stereotypes and prejudice. In J. F. Dovidio, M. Hewstone, P. Glick, & V. Esses (Eds.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 241–257). Thousand Oaks, CA: Sage.
Nadeau, S., Kataoka, N., Valerio, A., Neuman, M. J., & Elder, L. K. (2011). *Investing in young children: An Early Childhood Development Guide for policy dialogue and project preparation*. Washington: The World Bank.
Paluck, E. L., & Green, D. P. (2009). Prejudice reduction: what works? A review and assessment of research and practice. *Annual Review of Psychology, 60*, 339–367.
Patchen, M. (1982). *Black–White contact in schools: Its social and academic effects*. West Lafayette, IN: Purdue University Press.
Patterson, M. M., & Bigler, R. S. (2006). Preschool children's attention to environmental messages about group: Social categorization and the origins of intergroup bias. *Child Development, 77*, 847–860.
Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*, 751–783.
Pfeifer, J. H., Brown, C. S., & Juvonen, J. (2007). Teaching tolerance in schools: lessons learned since Brown vs Board of Education about development and reduction of children's prejudice. *Social Policy Report: Giving child and youth development knowledge away, 21*(2).
*Perkins, D. M., & Mebert, C. J. (2005). Efficacy of multicultural education for preschool children: A domain-specific approach. *Journal of Cross-Cultural Psychology, 36*, 497–512.
*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 Quarterly, 18*, 530–546.
Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer-Verlag.

Raabe, T., & Beelmann, A. (2011). Development of ethnic, racial, and national prejudice in childhood and adolescence. A multinational meta-analysis of age differences. *Child Development, 82*, 1715–1737.

*Rutland, A., Cameron, L., Bennett, L., & Ferrell, J. (2005). Interracial contact and racial constancy: A multi-site study of racial intergroup bias in 3–5 year old Anglo-British children. *Journal of Applied Developmental Psychology, 26*, 699–713.

Slater, M. D., & Rouner, D. (2002). Entertainment-education and elaboration likelihood: Understanding the processing of narrative persuasion. *Communication Theory, 12*, 173–191.

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

Tropp, L., & Prenovost, M. (2008). The role of intergroup contact in predicting children's inter-ethnic attitudes: Evidence from meta-analytic and field studies. In S. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood* (pp. 236–248). New York: Oxford University Press.

*Turner, R. N., Hewstone, M., & Voci, A. (2007). Reducing explicit and implicit outgroup prejudice via direct and extended contact: The mediating role of self-disclosure and intergroup anxiety. *Journal of Personality and Social Psychology, 93*, 369–388.

van Sluijs, E. M. F., McMinn, A. M., & Griffin, S. J. (2009). Effectiveness of interventions to promote physical activity in children and adolescents: systematic review of controlled trials. *British Medical Journal*. http://dx.doi.org/10.1136/bmj.39340.843947.BE.

*Verkuyten, M., & De Wolf, A. (2007). The development of in-group favoritism: Between social reality and group identity. *Developmental Psychology, 43*, 901–922.

*Wham, M. A., Barnhart, J., & Cook, G. (1996). Enhancing multicultural awareness through storybook reading experience. *Journal of Research and Development in Education, 30*, 1–9.

*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.