



Teaching tolerance or acting tolerant? Evaluating skills- and contact-based prejudice reduction interventions among Palestinian-Israeli and Jewish-Israeli youth

Alaina Brenick^{a,*}, Samantha E. Lawrence^b, Daniell Carvalho^b, Rony Berger^c

^a University of Connecticut, Department of Human Development and Family Sciences, 348 Mansfield Rd., U-1058, Storrs, CT 06269-1058, United States of America

^b University of Connecticut, United States of America

^c Ben Gurion University of the Negev, Israel

ARTICLE INFO

Mohini Lokhande

Keywords:

Intergroup contact
Intergroup social exclusion
Middle-east
Prejudice reduction interventions
Social-cognitive skills
Social domain model
Social-emotional skills

ABSTRACT

Although contact-based interventions are the cornerstone of prejudice reduction, in high-conflict environments, incendiary contact with outgroups can instead exacerbate negative attitudes. Supplementing contact interventions with social-cognitive/emotional approaches may, instead, help facilitate positive contact. This study evaluated the effectiveness of two prejudice reduction interventions among 148 Palestinian-Israeli and 154 Jewish-Israeli 5th grade students ($M_{age,years} = 10.55$, $SD = 0.26$) in a high conflict area. Schools in Jaffa, Israel were assigned to a social-cognitive/emotional skills-based intervention, a skills- and contact-based intervention (i.e., skills, skills+contact), or the control group—all delivered as part of the curriculum. Prejudice was assessed through participants' judgments of and justifications about hypothetical scenarios of intergroup exclusion in peer and home contexts at pre-test, post-test, and 6-month follow-up. Repeated measures ANOVAs showed various main effects including gender, ethnicity, and context in which the exclusion occurred (peer/home). Significant higher level interactions with group by time demonstrated the positive influence of both treatment groups on prejudice reduction. The skills and skills+contact groups became *more rejecting* while the control group became *more accepting* of exclusion across time. Additionally, the skills and skills+contact groups increased in moral and empathic reasoning over time, whereas the control group increased in social conventional and stereotyped prejudiced reasoning. These findings illustrate the effectiveness of in-school social-cognitive/emotional skills and combined skills+contact approaches in reducing the prejudiced attitudes of Palestinian- and Jewish-Israeli pre-adolescents, especially in areas with protracted conflict.

1. Introduction

Ethnic prejudice and discrimination are widespread social phenomena that adversely affect children's psychological development, safety, and overall well-being (Save the Children, 2006), leading to such consequences as health problems (Taylor, 2015), behavioral difficulties (Tobler et al., 2013), poor school achievement (Benner, Crosnoe, & Eccles, 2015), compromised emotional growth (Schmitt, Branscombe, Postmes, & Garcia, 2014), and social exclusion (Rutland & Killen, 2015). Of note, just as social exclusion may

* Corresponding author.

E-mail address: alaina.brenick@uconn.edu (A. Brenick).

<https://doi.org/10.1016/j.jsp.2019.07.001>

Received 8 March 2018; Received in revised form 31 January 2019; Accepted 2 July 2019

Available online 09 August 2019

0022-4405/ © 2019 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

be a consequence of ethnic prejudice, socially excluding others based on their group membership (group-based exclusion), is in and of itself a form of prejudice (Cooley, Elenbaas, & Killen, 2016). Moreover, this particular form of prejudice is a typical manifestation of the negative effects of living amidst protracted conflict in the Middle-East in the everyday lives of youth, even from a very young age (Alsamih & Tenenbaum, 2018; Bar-Tal & Teichman, 2005; Brenick et al., 2007, 2010).

Prejudice, stereotyping, and their associated detriments warrant particular concern in contexts of violent, protracted conflict because such negative intergroup attitudes can crystallize by adolescence (see Raabe & Beelmann, 2011; Rutland & Killen, 2015)—a critically polarizing experience for these youth (Bar-Tal, Diamond, & Nasie, 2017)—into conflict-promoting narratives that not only delegitimize the outgroup, but may also encourage the acceptance of forced segregation and violence directed toward the outgroup (Bar-Tal et al., 2017). As a result, it is of the utmost importance to intervene early and implement prejudice reduction interventions in pre-adolescent populations as a means of preventing the support for continued violent conflict.

The current study comprised a longitudinal analysis of school-based interventions, utilizing the most widely applied and effective approaches for prejudice reduction within the context of intergroup social exclusion: social-skills training and contact-based approaches (Aboud et al., 2012; Aboud & Levy, 2000; Beelmann & Heinemann, 2014; Lemmer & Wagner, 2015). More specifically, Palestinian-Israeli (P-I) and Jewish-Israeli (J-I) 5th grade students were divided into three groups: two treatment groups (i.e., skills, skills+contact) and one control group that received the standard social studies curriculum (i.e., control). Prejudiced intergroup attitudes were based on participants' evaluations of social exclusion scenarios in which an outgroup member was excluded from a peer- or family home-based activity and were assessed pre- and post-intervention, as well as at a 6-month follow-up. In what follows, we briefly describe the various theoretical approaches to assessing prejudiced social exclusion, as well as prejudice reduction and positive intergroup attitude promotion with children, outline the major intergroup programs that employ skills- and contact-based interventions, and provide a developmental rationale for this study.

1.1. Evaluating social exclusion: Social Domain Model

The Social Domain Model (Turiel, 1983; see Smetana, Jambon, & Ball, 2014) provides a theoretical and methodological framework for this study through which judgments about social issues, such as discriminatory group-based social exclusion, are systematically evaluated and understood in terms of the negotiation between three domains of social thought: *moral* considerations of justice, welfare, and equality; *societal* conventions, social expectations, and group norms; and *psychological* constructs of personal preference or choice. The manner in which these domains are coordinated when evaluating a social interaction (e.g., group-based social exclusion, GBE) is directly related to beliefs about the acceptability, or lack thereof, of that interaction and whether or not any action should be taken to address it (see Brenick & Halgunseth, 2017; Horn & Nucci, 2006; Nucci & Turiel, 2009; Palmer & Abbott, 2017).

Studies conducted in diverse, but peaceful, regions have found that when children deem exclusion based on ethnicity to be wrong, they cite moral concerns, but when they accept discriminatory group-based exclusion (GBE), they do so based on stereotyped beliefs about groups and group dynamics (Abrams, Palmer, Rutland, Cameron, & Van de Vyver, 2014; Alsamih & Tenenbaum, 2018; Brenick & Killen, 2014; Brenick & Killen, 2014; Malti, Killen, & Gasser, 2012; Tenenbaum, Leman, & Aznar, 2017). Moreover, youth look to their salient groups, such as peers and parents/family (Rutland, Killen, & Abrams, 2010), for (perceived) norms about interpreting intergroup interactions; they also tend to be more accepting of exclusion that occurs within the realm of the family (e.g., the family home; Brenick & Romano, 2016). However, only a handful of studies have assessed such reasoning in areas of violent or extended intergroup conflict (Brenick et al., 2007, 2010), where social exclusion of outgroup members might not only be a typical experience, but also one that is *encouraged* by group, family, and societal norms (see Alsamih & Tenenbaum, 2018; Brenick & Romano, 2016).

Therefore, we aimed to explore the role of context, characterized by the aforementioned socio-cultural norms about intergroup attitudes and deference to authority (i.e., are pre-adolescents differentially influenced by and deferent to peers versus families?), in pre-adolescents' manifestations of prejudice. Moreover, no intervention studies have yet assessed children's reasoning about discriminatory GBE, let alone in areas of conflict where such interventions are so desperately needed. Thus, the current study pulls from the broader field of prejudice reduction interventions to evaluate the effectiveness of two interventions in reducing the negative intergroup attitudes of P-I and J-I pre-adolescents from several schools in a high conflict community.

1.2. Prejudice reduction interventions

Research efforts to prevent or reduce negative intergroup attitudes have led to the identification of numerous individual and social influences on prejudice and stereotyping, ranging from intergroup contact (Tropp & Prenovost, 2008) to classification abilities (Bigler & Liben, 2007), and many more (e.g., Killen & Rutland, 2011; Nesdale, 2004; Rutland et al., 2010; Smetana, 2006). However, skills- and contact-based intervention approaches have been found to be the most effective approaches to reduce prejudice (Beelmann & Heinemann, 2014) and, thus, form the basis of the approaches analyzed in the current study.

1.2.1. Skills-based interventions

Social-cognitive and social-emotional developmental approaches to prejudice reduction suggest that—especially beginning in childhood—prejudice reduction programs must account for and address the social-cognitive and social-emotional skills underlying intergroup dynamics, including social categorization, perspective-taking, empathy, and moral reasoning (e.g., Berger, Benatov, Abu-Raiya, & Tadmor, 2016; Bigler & Liben, 2007; Killen & Rutland, 2011; Nesdale, 2004). These interventions are rooted in developmental theories that emphasize the influence of shifts in reasoning and burgeoning social-cognitive skills (Aboud, 2008; Bigler &

Liben, 2007; Rutland & Killen, 2015), as well as increases in social and self-awareness, relationship skills, and responsible decision-making, among other social-emotional learning competencies (Malti & Noam, 2016) on children's outgroup prejudice.

1.2.2. Contact-based interventions

Contact-based interventions have been the most widely applied, studied, and consistently effective prejudice reduction approach (Aboud et al., 2012; Beelmann & Heinemann, 2014; Pettigrew & Tropp, 2006). These interventions draw upon contact theory (Allport, 1954), which asserts that, if people engage in meaningful cross-group interactions (i.e., contact that involves the “optimal” conditions of equal status, common goals, intergroup cooperation, and support of authorities), they are more likely to understand and accept each other and show reduced prejudice as a result. Researchers later asserted the importance of time to develop intergroup friendships as a facilitating condition for effective intergroup contact (Davies, Tropp, Aron, Pettigrew, & Wright, 2011; Pettigrew, 1998).

Contact-based approaches employed with school children have utilized many strategies, such as: integrated structured social activities (Berger, Benatov, et al., 2016), cooperative learning (Aronson & Patnoe, 1997), use of media (e.g., TV, books) depicting contact (Brenick et al., 2007; Cameron & Rutland, 2006), and bilingual education (Bekerman, 2005). Overall, there is ample evidence for the effectiveness of intergroup contact in reducing prejudice and stereotypes (Jugert & Feddes, 2017; Pettigrew & Tropp, 2006; Titzmann, Brenick, & Silbereisen, 2015) and promoting prosocial moral reasoning about intergroup relations (Brenick & Killen, 2014; Ruck, Park, Crystal, & Killen, 2015) both in peaceful multicultural societies and conflict zones around the world (see Berger, Abu-Raiya, & Gelkopf, 2015; Berger, Benatov, et al., 2016; Berger, Brenick, & Tarrasch, 2018; Brenick et al., 2007; Lemmer & Wagner, 2015).

Nonetheless, conducting contact-based interventions with members of highly conflict plagued groups *can* be challenging and, at times, incendiary (Barlow et al., 2012; Guffler & Wagner, 2017). Thus, the current study evaluated the efficacy of a combined skills- and contact-based approach to determine if incorporating both components might prevent some of the potentially exacerbating effects of a contact-based intervention in a region of protracted conflict.

1.3. Context of the current study

This study was conducted in Jaffa, Israel—the oldest part of the Tel Aviv-Yafo municipality. Historically, the area has been populated by P-Is; however, more recently J-Is have migrated to the region. Approximately 20,000 P-Is and 35,000 J-Is reside in this town, although the two populations remain largely segregated. P-I and J-I children attend separate public schools run by the Israeli Ministry of Education. In addition to their regular curricula, public schools in the Jewish sector teach in Hebrew and offer Jewish history, religion, and culture, whereas public schools in the Palestinian sector teach in Arabic and offer lessons in Arab history, religion, and culture. In recent years, tensions between the P–I and J–I populations have risen due to the escalation of violence in the P–I conflict (Shor, 2017). A wealth of research has shown that the ongoing conflict manifests in the day-to-day lives of P–I and J–I youth through outgroup prejudices, unwillingness to engage with the outgroup, and acceptance of outgroup social exclusion (e.g., Bar-Tal & Teichman, 2005; Berger et al., 2018; Brenick et al., 2007, 2010), which can develop into more polarizing discrimination and violence later in adolescence and into adulthood (Bar-Tal et al., 2017).

With P–I and J–I youth growing up amidst the ongoing conflict, deeply rooted historical memories of (Bar-Tal, 2011, 2013), the everyday reality of (e.g., military operations, terror attacks), and societal beliefs based on the ethos of conflict (see Brenick & Romano, 2016) permeate their social identities. Thus, there is a risk for negative interaction between group members that can lead to inflammatory, rather than tolerance building, effects. In fact, some scholars caution that when groups have a long history of intergroup animosity and hostility, negative contact may lead to more prejudice against the outgroup (Brown & Hewstone, 2005; Maoz, 2011; Salomon, 2006; Vorauer & Sasaki, 2011). However, the positive results from various contact-based interventions in conflict zones (e.g., Berger, Benatov, et al., 2016; Brenick et al., 2007; Lemmer & Wagner, 2015), coupled with the finding that intergroup friendships are the best predictor of prejudice reduction (Tropp & Prenovost, 2008), beg the question of how contact interventions can be effectively implemented under trying conditions with consistently positive results.

1.4. The current study

The current study evaluated the differential impacts of two treatments on P–I and J–I youths' attitudes about the exclusion of outgroup members: the individual intervention of social-cognitive and -emotional skills training and a combined skills training and intergroup contact intervention. Participants were divided into three groups: two treatment groups (i.e., skills group and skills + contact group) and one control group (i.e., social studies group). Prejudice was assessed based on participants' judgments (i.e., how good or bad is the outgroup exclusion?) and justifications (i.e., why is it good or bad?) of social exclusion scenarios, in which an outgroup member was excluded from a peer or home context (i.e., a soccer game or a sleepover in the family home, respectively), which were measured pre- and post-intervention, as well as at a 6-month follow-up.

We hypothesized that participants in the treatment groups, as compared to participants in the control group, would show 1) significant reductions in their acceptance of outgroup social exclusion and their use of stereotyped reasons to justify that acceptance and 2) significant increases in rejecting outgroup social exclusion, accompanied by increases in their appeals to moral and empathic concerns about exclusion. Further, given pre-adolescents' likelihood to be more accepting of exclusion from the home or due to parental norms of outgroup exclusion (see Alsamir & Tenenbaum, 2018; Brenick & Romano, 2016), we hypothesized that participants would be more accepting of 3) parent-sanctioned GBE as compared to peer-sanctioned or unspecified group-based exclusion,

and 4) exclusion from the family home than from a peer activity outside of the home. Additionally, because previous research has not yet compared a skills and additive skills+contact intervention, particularly in the context of protracted conflict, it was an open question as to whether 5) social-emotional/cognitive skills training can complement and ensure the effectiveness of intergroup contact, or if the skills and skills+contact interventions might be differentially effective in reducing prejudice and promoting positive intergroup attitudes, and 6) whether the interventions' effectiveness might also differ by ethnic group.

2. Method

2.1. Setting

This study was conducted between September 2015 and December 2016 in Jaffa. The department of education in the municipality invited the Arab-Jewish Community Center to develop a program that would diffuse tension and create a cultural partnership between P–I and J-I students in Jaffa. The program was approved by the Tel Aviv University Institutional Review Board and the education department of the municipality of Tel Aviv-Jaffa's ethical committee. The latter granted access to public schools, pending local approval from the schools. Three P–I and three J-I elementary schools in Jaffa were selected as sites for the study due to their representativeness of schools in the area and their similar socio-economic indices (i.e., the ministry of education's index comprised of parents' education, per capita income, school periphery, and country of origin).

2.2. Participants

Participants were 302 P–I ($n = 148$) and J-I ($n = 154$) 5th grade students ($M_{age, years} = 10.55$, $SD = 0.26$). Classes at each of the schools were assigned, by coin toss, to either the social studies control group (control; $n = 97$), skills-oriented training intervention (skills; $n = 103$), or the skills- and contact-based intervention (skills+contact; $n = 102$). Participants whose parents signed consent forms for them to participate also assented to participate (only one parent did not provide consent). Participants were approximately evenly divided by ethnicity and gender (see Table 1 for full demographic breakdown). Six students were dropped from the study for failing to complete the questionnaires. No significant differences emerged at baseline between those who completed the study and those who dropped out.

2.3. Procedure for implementation of the interventions

Following principals' approval, the program's broader rationale was introduced to each of the twelve participating classes' homeroom teachers (two classes per school). Thereafter, the teachers were instructed on their group assignments. The skills and the control group teachers were instructed to prepare the students for the programs and remind them to practice the learned skills between sessions. The skills+contact group teachers were instructed to accompany the students during the Arab-Jewish Community Center's meetings and serve as observers rather than play an active role in facilitation. All teachers were asked to present the program to the children's parents and encourage their support.

The facilitators for all groups were bilingual P–I and J-I graduate students recruited based on their experience working with multicultural youth groups. Prior to the intervention implementation, the facilitators underwent a 6-hour training led by the fourth author in order to ensure adherence to the intervention.

2.3.1. The social studies control group

The social studies control group consisted of twelve 45-minute bi-monthly sessions. The intervention was derived from the Key to

Table 1
Breakdown of participant gender and ethnicity by treatment group.

Treatment Group	Ethnicity	Gender		Total
		Female	Male	
Control	Palestinian-Israeli	23	25	48
	Jewish-Israeli	29	40	69
	Total	52	45	97
Skills	Palestinian-Israeli	26	23	49
	Jewish-Israeli	28	26	54
	Total	54	49	103
Skills + Contact	Palestinian-Israeli	25	26	51
	Jewish-Israeli	27	24	51
	Total	52	50	102
Total	Palestinian-Israeli	74	74	148
	Jewish-Israeli	84	70	154
	Total	158	144	302

the Heart social studies curriculum, which aims to cultivate students' civic values, to nurture relationship between the citizen and the society, to construct a framework of rules and procedures for social-ethical learning and to encourage social and ethical discourse (Interagency Task Force on Israeli Arab Issues, 2016). The facilitator delivered monthly thematic modules during regularly scheduled social studies classes. Module themes included: sharing and participating, social and community involvement, understanding diversity, conflict resolution, promoting respect, accepting the “other,” and facilitating a safe and secure school atmosphere. These topics were taught via lectures, stories, and experiential exercises.

2.3.2. The skills training intervention

Whereas the control group fostered general civic values and prosocial behaviors, the skills intervention focused on practicing specific skills related to developing non-judgmental attitudes, cognitive and emotional empathy, and compassion for self and others (see online supplement A for more detailed intervention information). The skills intervention consisted of twelve 45-minute bi-monthly sessions that combined perspective-taking and empathy training with developmentally appropriate contemplative practices (i.e., mindfulness and compassion meditation). It borrowed practices from three sources: Enhancing Resiliency Among Students Experiencing Stress and Promoting Pro-Social Behavior, a social-emotional program (Berger, Gelkopf, Heineberg, & Zimbardo, 2016); perspective-taking and empathy training (Doyle & Aboud, 1995; Frey, Nolen, Van Schoiack Edstrom, & Hirschstein, 2005); and the Call to Care, a mindfulness and compassion-cultivating program (Berger et al., 2018).

Each session started with a psycho-educational presentation of the session's theme, followed by an explanation or demonstration of the skills to be practiced by the students. The bulk of the session was devoted to practicing the skills and discussing experiences that resulted from the practice. Students were encouraged to share with their parents the skills learned in these sessions and to practice them in between sessions. Following each session, the students' parents were informed of the materials learned in the session via the school's website or email and were encouraged to discuss and, if possible, practice the skills with their children.

2.3.3. The skills + contact intervention

Due to the potentially incendiary nature of negative contact (Brown & Hewstone, 2005; Salomon, 2006; Vorauer & Sasaki, 2011), the current study did not include a contact-only intervention group. Instead, it had a skills + contact intervention which, in addition to the skills group activities listed above, consisted of twelve bi-monthly meetings of ethnically mixed students. The aim of these meetings was to familiarize the students with each other, while highlighting their similarities and differences (Dovidio, Gaertner, & Saguy, 2007). For the majority of the contact portion of the intervention, students were divided into mixed ethnic groups of 15–18 students and, in each 4-hour session, the mixed groups engaged in warm-up exercises, experiential work related to the theme of the session, group discussions, and they rotated between three 1-hour artistic activities (e.g., music activity, movement, social play). In between these activities, the students had a half-hour break where they played and had snacks.

To meet the optimal conditions originally suggested by Allport (1954), the following was done: first, the sessions were conducted with a similar number of P–I and J–I students, both languages were used interchangeably, and both P–I and J–I facilitators were present in all activities (i.e., equal status). Although the sessions were conducted in both Hebrew and Arabic to ensure that the two cultures were on equal standing, the P–I and the J–I students were all fluent in Hebrew and able to communicate with one another. Second, all activities were geared to achieve goals desirable to the students and required face-to-face interactions and cooperation (i.e., common goals and intergroup cooperation). Third, the program was sponsored and supported by the municipality's education department, the school administration, and the parents (i.e., support of authorities). Finally, the longer duration of the sessions provided time to develop greater intimacy (i.e., facilitated friendship building) and the encouragement of the facilitators and the homeroom teachers fostered continued contact between the students during and after the program (e.g., encouraging them to exchange phone numbers and email addresses; see online supplement B for more detailed intervention information).

2.4. Fidelity of the interventions

Two trained research assistants observed and rated the fidelity of all group program implementations. Assessments on three domains—1) facilitators covered the intended topics; 2) facilitators followed the practices or experiential activities; and 3) group members were active during the session—were calculated on a 6-point scale ranging from 0 (*not at all as stipulated in the protocol*) to 5 (*exactly as stipulated in the protocol*). Fidelity was acceptable for all groups with good to high interrater agreement. The average fidelity ratings were as follows: 1) control group: $M = 3.20$, $SD = 0.82$, (Cohen's $K = 0.65$, $p < 0.01$); 2) skills group: $M = 4.60$, $SD = 0.46$, ($K = 0.87$, $p < 0.001$); and 3) skills + contact group: $M = 4.10$, $SD = 0.61$, ($K = 0.81$, $p < 0.001$). The ratings of fidelity did not differ significantly by group (all p 's > 0.17).

2.5. Procedure for evaluating the interventions

Five outcome indicators were used to measure prejudiced evaluations of intergroup exclusion. Participants were assessed a week before the intervention began, immediately after it was completed, and at a 6-month follow-up. The questionnaires were administered in written format by two trained research assistants who spoke Arabic and Hebrew and were able to assist the students as needed; they were blind to the group condition of the students.

2.6. Measure: Evaluations about Intergroup Exclusion Survey

Prejudiced intergroup attitudes were assessed using the Evaluations about Intergroup Exclusion Survey, which included vignettes depicting intergroup exclusion in a peer (e.g., social activity with a group of friends') and a home (e.g., group sleepover in the family home) scenario. Multiple versions of the same vignettes were created so that participants only read vignettes in which the protagonist was of the same gender and ethnicity as them. Participants evaluated five items per context. The survey was translated and back translated between Arabic and Hebrew and pilot-tested by children and expert adults fluent in both languages to ensure ecological validity and equivalence of the scenarios. An assessment of measurement invariance (Fischer & Fontaine, 2011) found the measure equivalent across the two groups ($\phi(P-I, J-I) = 0.99$; Lorenzo-Seva & ten Berge, 2006). The measure was also highly reliable at all time points for both groups (Cronbach's alphas P-I: T1 = 0.94, T2 = 0.95, T3 = 0.94; J-I: T1 = 0.96, T2 = 0.96, T3 = 0.96).

First, participants read exclusion vignettes for each scenario and rated how good or bad the exclusion was based on a number of criteria. As an example, the peer exclusion scenario for J-I male participants described a J-I protagonist who, with his friends, was organizing a soccer game. There was a P-I outgroup member (X) and a J-I ingroup member (Y) who both wanted to join the game. Both were good soccer players, but there was room for only one more on the team. The protagonist invited the ingroup member. Upon finishing reading the vignette, participants were then asked, "How good or bad is it to exclude X and include Y?": 1) when no reason was specified (undifferentiated exclusion), 2) "...because X is Palestinian?" (group-based exclusion GBE, undefined), 3) "... because [the protagonist's] parents say to exclude X because he is Palestinian?" (GBE, parent-sanctioned), 4) "... because [the protagonist's] friends say to exclude X because he is Palestinian?" (GBE, peer-sanctioned). Judgments were rated on a 7-point Likert-type scale (1, very, very good, to 7, very, very bad), with lower numbers indicating more prejudiced attitudes accepting of discriminatory intergroup exclusion.

To assess how participants made sense of their attitudes about intergroup exclusion, they were also asked to provide justifications for their ratings; they were asked, 5) "Why is it good or bad if the protagonist and his friends don't let X join the game?" A series of eight possible justifications, divided into five categories, were provided for participants to select from, and they were able to select up to three responses for why they thought the exclusion was good or bad. Participants' selection(s) of justifications from the five categories were coded as a proportion of one. If only one justification category was selected, it was coded as 1; if two were selected, they were each coded as 0.50; if three were selected, they were each coded as 0.33. If a justification was not selected, it was coded as zero.

Drawing on Social Domain Theory (Turiel, 1983), justifications were categorized into 1) societal, 2) moral, and 3) psychological domains. The societal domain (social conventional) included justifications such as, "they wouldn't be friends because X is Palestinian," and "X wouldn't fit in with them." Concerns for about harm and welfare, such as how "mean it is" and how "it might hurt X's feelings," for example, characterized the moral domain (moral). The justification that it is the protagonist's own personal choice whom they include or exclude, represented the psychological domain (personal choice). Furthermore, based on previous empirical studies from the Social Reasoning Developmental Perspective (Rutland et al., 2010), we also included 4) stereotyped conceptualizations of the outgroup (stereotype) as a distinct sub-domain ("the outgroup kid probably did something to deserve it."). Additionally, because our interventions had the potential to increase empathy overall—not just toward the excluded individuals, but also toward the excluders—we included a separate indicator of the moral domain—5) empathy directed toward the *excluder*, not the excluded (*excluder* empathy).

3. Results

3.1. Plan for analyses

First, we evaluated the intraclass correlation coefficients (ICC) to determine whether administering the interventions at the classroom level warranted multilevel analysis of the data due to significant clustering effects. The analyses yielded ICCs ranging from 0.00 to 0.23 with a low mean of 0.03 for justifications and an average estimated sample size of 302, indicating adequate power for our study (online supplement Table 1a). Higher ICCs were found only for the judgment dependent variables ($M = 0.19$). Therefore, we carried out supplemental analyses to assess the potential impact of classroom clustering on the success of the intervention on participants' judgments. Linear regression models were carried out for treatment group effects predicting participants' judgment ratings at each wave, while controlling for ethnicity and gender. We then controlled for classroom clustering error bias in the treatment group effect standard errors with wild cluster bootstrapping (see Cameron, Gelbach, & Miller, 2008) using the multiwaycov package for R (Graham, Mahmood, & Björn, 2016). Overall, the resulting bootstrapped confidence intervals remained similar or decreased—indicating robustness of the findings presented below, though there were a handful of instances in which they increased, (see online supplement Tables 2a–5a). The wild cluster bootstrapping results, along with the lack of clustering effect on justifications, and the small number of clusters, warranted repeated measures ANOVAs—as opposed to multilevel modeling or correcting the standard errors—as the most suitable method for carrying out the main analyses.

Since, overall, the results of these analyses indicated low levels of variance accounted for by clustering, a multivariate ANOVA was conducted to assess whether pre-test levels of each outcome variable differed significantly by group or ethnicity. No significant group differences emerged for any outcome variable other than the excluder empathy justification. However, significant differences emerged for ethnicity on almost every outcome variable (see online supplement Table 6a for all results). To account for these significant baseline differences, ethnicity was treated as a covariate in all further analyses.

Next, a series of repeated measures ANOVAs were carried out to assess our main research questions regarding how the interventions might differentially relate to participants' judgments of and justifications about scenarios of intergroup social exclusion across time. To assess judgments of undifferentiated exclusion, a 3 (group: control, skills, skills + contact) x 2 (scenario: peer, home) x

Table 2
Results of repeated measures ANOVA for undifferentiated exclusion judgments.

	Sum of Squares	df	Mean Square	F	η_p^2	p
Significant within-subjects effects						
Scenario	68.43	1	68.43	37.35	0.11	***
Scenario X Ethnicity	15.49	1	15.49	8.46	0.03	**
Error (Scenario)	544.19	297	1.83			
Time X Group	31.29	4	7.82	12.36	0.08	***
Error (Time)	376.06	470	0.80			
Significant between-subjects effects						
Group	72.71	2	36.35	4.22	0.03	*
Ethnicity	842.54	1	842.54	97.79	0.25	***
Gender	168.42	1	168.42	19.547	0.06	***
Error	2558.89	297	8.62			

Note. Only significant effects are reported.

* = $p < 0.05$.

** = $p < 0.01$.

*** = $p < 0.001$.

3 (time: pre-test, post-test, follow-up) ANOVA was conducted with repeated measures on the last two variables. Participants' judgments of group-based exclusion (GBE), as sanctioned by parents, peers, or an undefined source, were assessed with a 3 (group) x 2 (scenario) x 3 (time) x 3 (source: undefined, parent-sanctioned, peer-sanctioned) ANOVA with repeated measures on the last three variables. Finally, participants' justifications of their judgments regarding GBE were assessed with a 3 (group) x 2 (scenario) x 3 (time) x 3 (source: undefined, parent sanctioned, peer-sanctioned) x 5 (justification: social conventional, stereotype, personal choice, moral, excluder empathy) ANOVA with repeated measures on the last four variables.

Gender (2: male, female) and ethnicity (2: P–I, J–I) were included as covariates for all analyses. In cases where the sphericity assumption was not met, the Huynh-Feldt correction was used. Significant main effects were followed-up with pairwise comparisons. Significant interactions including within subjects variables were explored with post-hoc repeated measures and univariate ANOVAs. Bonferroni adjustments were used to control for multiple comparisons. In order to provide a metric for the group differences, Cohen's (1988) d indices of effect size were computed, comparing differences across the three time points between the skills, skills + contact, and control groups, using pooled difference standard deviations. Missing data were minimal in this study ranging from 0 to 0.6%. Participants were excluded (pairwise) only from specific analyses in which their data were missing.

3.2. Undifferentiated exclusion judgments

The analyses of participants' undifferentiated exclusion judgments yielded significant main effects for gender, scenario, ethnicity, and group. The effects for scenario, ethnicity, and group, however, were qualified by significant scenario by ethnicity and time by group interactions (Table 2).

The main effect for group, which was qualified by a time by group interaction, revealed group-specific patterns of change in judgments of undifferentiated intergroup exclusion over the three time points (Fig. 1). The skills group became significantly more rejecting of exclusion from pre-test to post-test, and this effect was maintained at follow-up ($F(2, 204) = 7.22, p < 0.01, \eta_p^2 = 0.07$). Likewise, the skills + contact group became more rejecting of exclusion from pre-test to post-test and then to the follow-up. However, only ratings at the pre-test and follow-up differed significantly from one another ($F(2, 202) = 4.98, p < 0.01, \eta_p^2 = 0.05$). In contrast to the intervention groups, the control group became significantly more accepting of undifferentiated exclusion of the outgroup at each time point ($F(2, 192) = 14.55, p < 0.001, \eta_p^2 = 0.13$). Cohen's d values comparing the differences between the two treatment groups and the control group averages yielded small to moderate effect sizes. There were no differences in effect sizes between the two treatment groups (see Table 3 for all effect sizes).

3.3. Group-based exclusion (GBE) judgments

The repeated measures ANOVA for evaluations of GBE yielded several significant main effects: gender, scenario, source, group, and ethnicity. The main effects for source, group, and ethnicity, however, were qualified by significant source by time, time by group, source by time by group, and source by ethnicity interactions (Table 4).

The source main effect demonstrated that participants were most accepting of parent-sanctioned ($M = 3.45, SD = 1.61$) and least accepting of undefined ($M = 4.46, SD = 1.44$) GBE with ratings of peer-sanctioned GBE falling in between ($M = 3.79, SD = 1.57$)—all three ratings were significantly different from one another at $p < 0.001$. This main effect and the source by time and time by group interactions need to be interpreted through the three-way interaction between source, time, and group. First, follow-up analyses of this three-way interaction revealed that, when the source of GBE was undefined, participants in the skills ($F(2, 204) = 21.45, p < 0.001, \eta_p^2 = 0.17$), and skills + contact groups ($F(2, 202) = 12.84, p < 0.001, \eta_p^2 = 0.11$), became significantly more rejecting of outgroup exclusion from pre-test to post-test, and these effects were maintained at follow-up. In contrast, the control group became significantly more accepting of undefined GBE from pre-test to post-test, and then from post-test to 6-month

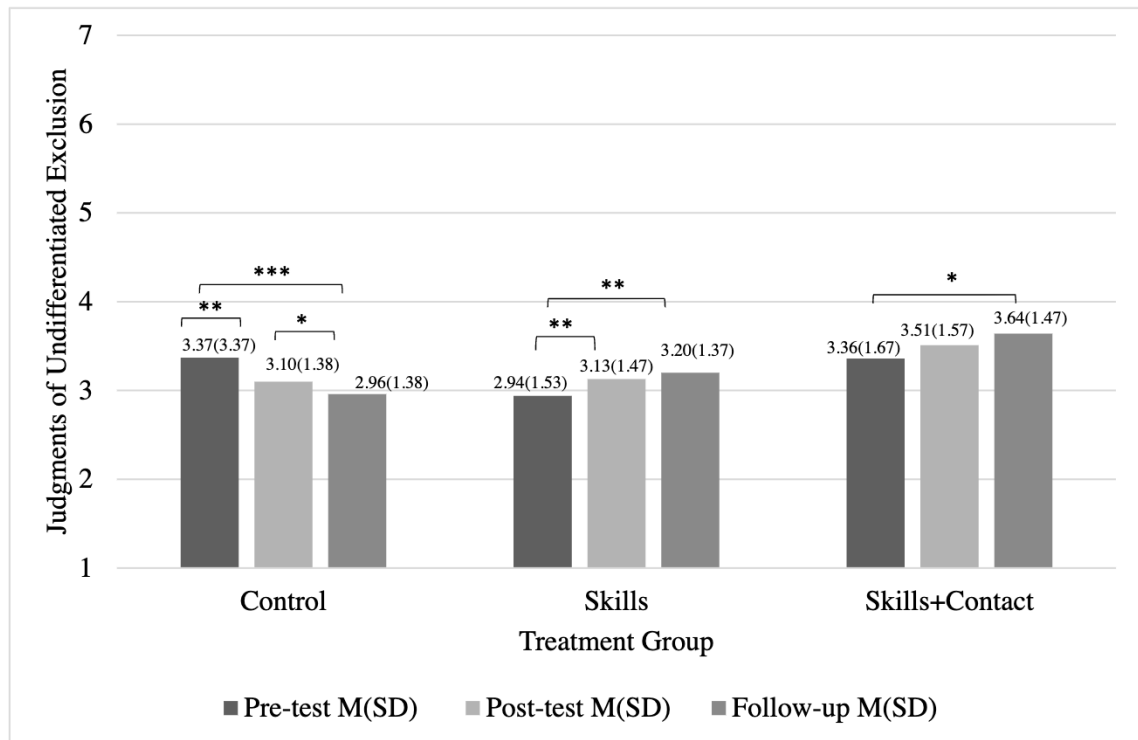


Fig. 1. Interaction Between Time and Group for Undifferentiated Exclusion Judgments

Note. Bars represent mean judgments of undifferentiated exclusion; Judgments were rated on a 7-point Likert-type scale (1, very, very good, to 7, very, very bad); Brackets indicate significant differences in judgments across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

follow-up ($F(2, 192) = 22.10$, $p < 0.001$, $\eta_p^2 = 0.19$; Fig. 2a). Cohen's d values assessing the magnitude of effect of the treatment groups as compared to the control group, yielded moderate pre- to post-test and large pre-test to follow-up effect sizes. There were no differences in effect sizes between the two treatment groups (see Table 3).

Second, when the source of GBE was explicitly defined as parents (parent-sanctioned), the skills ($F(2, 204) = 19.42$, $p < 0.001$, $\eta_p^2 = 0.16$) and the skills + contact ($F(2, 192) = 20.40$, $p < 0.001$, $\eta_p^2 = 0.18$) groups became more rejecting of GBE from pre-test to post-test, and this effect was maintained at follow-up. However, the control group grew increasingly accepting of parent-sanctioned exclusion from pre-test, to post-test, and to follow-up ($F(2, 192) = 16.44$, $p < 0.001$, $\eta_p^2 = 0.15$; Fig. 2b). Cohen's d values assessing the magnitude of effect of the treatment groups as compared to the control group, yielded moderate effect sizes from pre- to post-test and follow-up for the skills group, but very small effect sizes for the skills + contact group. This difference yielded small effect sizes when comparing between the two treatment groups—the magnitude of effect of the skills group was slightly larger than for the skills + contact group (see Table 3).

Third, when GBE was sanctioned by peers, both the skills ($F(2, 204) = 22.40$, $p < 0.001$, $\eta_p^2 = 0.18$) and skills + contact ($F(2, 202) = 7.48$, $p < 0.01$, $\eta_p^2 = 0.07$) groups became significantly less accepting of GBE from pre-test to 6-month follow-up. Specifically, the skills group became significantly more rejecting of peer-sanctioned GBE from pre-test to post-test, and this effect was maintained at follow-up, whereas the skills + contact group differed significantly only from pre-test to follow-up ($F(2, 202) = 5.68$, $p < 0.01$, $\eta_p^2 = 0.05$). In contrast, the control group became significantly more accepting of peer-sanctioned GBE from pre-test to post-test, and post-test to follow-up ($F(2, 192) = 16.44$, $p < 0.001$, $\eta_p^2 = 0.15$; Fig. 2c). Cohen's d values assessing the magnitude of effect of the treatment groups compared to the control group, yielded small to moderate effect sizes from pre- to post-test and large effect sizes from pre-test to follow-up. There were no differences in effect sizes between the two treatment groups (see Table 3).

3.4. Justifications for exclusion judgments

The repeated measures ANOVA for the five justification types (i.e., social conventional, stereotype, personal choice, moral, and excluder empathy) yielded significant main effects for: time, ethnicity, group, and justification. These main effects were all qualified by several interactions: justification by ethnicity, justification by gender, scenario by justification, justification by group (online supplement Table 7a), time by justification, and time by justification by group (Table 5).

The significant justification by group and justification by time interactions were qualified by a justification by time by group interaction. First, both the skills and skills + contact groups significantly decreased in their use of social-conventional (Fig. 3a) and stereotype (Fig. 3b) justifications from pre-test to post-test, and these effects were maintained at the follow-up. In contrast, both

Table 3Cohen's d effect size calculations across time and group for judgments and justifications.

Judgment by Treatment Group	d_{korr} pre- to post-test (Compared to Control)	d_{korr} pre- to follow-up (Compared to Control)	d_{korr} post- to follow-up (Compared to Control)	d_{korr} pre- to post-test (Compared to Skills)	d_{korr} pre- to follow-up (Compared to Skills)	d_{korr} post- to follow-up (Compared to Skills)	d_{korr} pre- to post-test (Compared to S + C)	d_{korr} pre- to follow-up (Compared to S + C)	d_{korr} post- to follow-up (Compared to S + C)
Judgments of undifferentiated exclusion									
Control	–	–	–	0.19	0.34	0.16	0.28	0.48	0.32
Skills	0.19	0.34	0.16	–	–	–	–0.01	0.05	0.06
S + C	0.28	0.48	0.32	–0.01	0.05	0.06	–	–	–
Judgments of group-based exclusion									
Averaged									
Control	–	–	–	0.49	0.72	0.23	0.38	0.61	0.23
Skills	0.51	0.70	0.19	–	–	–	–0.10	–0.04	0.07
S + C	0.38	0.61	0.23	–0.08	–0.06	0.03	–	–	–
Undefined GBE									
Control	–	–	–	0.58	0.77	0.19	0.48	0.70	0.22
Skills	0.58	0.77	0.19	–	–	–	–0.06	–0.01	0.05
S + C	0.48	0.70	0.22	–0.06	–0.01	0.05	–	–	–
Parent-Sanctioned GBE									
Control	–	–	–	0.43	0.56	0.13	0.16	0.18	0.01
Skills	0.43	0.56	0.13	–	–	–	–0.24	–0.34	–0.10
S + C	0.16	0.18	0.01	–0.24	–0.34	–0.10	–	–	–
Peer-Sanctioned GBE									
Control	–	–	–	0.48	0.71	0.22	0.34	0.61	0.26
Skills	0.48	0.71	0.22	–	–	–	–0.11	–0.06	0.06
S + C	0.34	0.61	0.26	–0.11	–0.06	0.06	–	–	–

Effect Size Calculations for Participants' Justifications of Exclusion

Treatment Group by Justification	d_{korr} pre- to post-test (Compared to Control)	d_{korr} pre- to follow-up (Compared to Control)	d_{korr} post- to follow-up (Compared to Control)	d_{korr} pre- to post-test (Compared to Skills)	d_{korr} pre- to follow-up (Compared to Skills)	d_{korr} post- to follow-up (Compared to Skills)	d_{korr} pre- to post-test (Compared to S + C)	d_{korr} pre- to follow-up (Compared to S + C)	d_{korr} post- to follow-up (Compared to S + C)
Control									
Empathy	–	–	–	0.60	0.64	0.04	0.47	0.54	0.07
Moral	–	–	–	0.98	0.92	–0.07	0.63	0.72	0.09
Stereotype	–	–	–	–0.83	–0.99	–0.15	–0.62	–0.70	–0.09
Per. Choice	–	–	–	–	–	–	–	–	–
Soc.Conv.	–	–	–	–0.98	–1.13	–0.15	–0.69	–0.85	–0.20
Skills									
Empathy	0.60	0.64	0.04	–	–	–	–0.14	–0.11	0.03
Moral	0.98	0.92	–0.07	–	–	–	–0.31	–0.28	0.03
Stereotype	–0.83	–0.99	–0.15	–	–	–	0.20	0.26	0.05
Per. Choice	–	–	–	–	–	–	–	–	–
Soc.Conv.	–0.98	–1.13	–0.15	–	–	–	0.29	0.29	–0.05
S ± C									
Empathy	0.47	0.54	0.07	–0.14	–0.11	0.03	–	–	–
Moral	0.63	0.72	0.09	–0.31	–0.28	0.03	–	–	–
Stereotype	–0.62	–0.70	–0.09	0.20	0.26	0.05	–	–	–
Per. Choice	–	–	–	–	–	–	–	–	–
Soc.Conv.	–0.69	–0.85	–0.20	0.29	0.29	–0.05	–	–	–

Note. Klauer's d_{korr} (2001; Morris, 2008); Control = Social Studies Control Group; Skills = Skills Treatment Group; S + C = Combined Skills and Contact Treatment Group; Per. Choice = Personal Choice Justification; Soc.Conv. = Social Conventional Justification.

groups used justifications appealing to moral reasoning (skills: $F(2, 204) = 50.37, p < 0.001, \eta_p^2 = 0.33$; skills+contact: $F(2, 202) = 16.47, p < 0.001, \eta_p^2 = 0.14$; Fig. 3c) and excluder empathy (skills: $F(2, 204) = 10.88, p < 0.001, \eta_p^2 = 0.10$; skills+contact: $F(2, 202) = 4.95, p < 0.01, \eta_p^2 = 0.05$; Fig. 3d) significantly more from pre-test to post-test. Increases in moral justifications were maintained at follow-up for both treatment groups, whereas increases in justifications based on excluder empathy were maintained at the follow-up only for the skills group. Neither treatment group differed significantly in their use of personal choice reasoning across the three time points (Fig. 3e).

Table 4
Results of repeated measures ANOVA for group-based exclusion judgments.

	Sum of Squares	df	Mean Square	F	η_p^2	p
Significant within-subjects effects						
Scenario	324.86	1	324.86	87.55	0.23	***
Error (Scenario)	1102.10	297	3.71			
Source	410.62	2	211.29	174.95	0.37	***
Source X Ethnicity	15.23	2	7.84	6.49	0.21	**
Error (Source)	697.09	577	1.21			
Time X Group	214.93	3	68.43	22.17	0.13	***
Error (Time)	1439.58	466	3.09			
Source X Time	4.63	3	1.43	6.46	0.02	***
Source X Time X Group	4.04	7	0.62	2.82	0.02	**
Error (Source X Time)	212.81	963	0.22			
Significant between-subjects effects						
Group	194.35	2	97.18	3.44	0.02	*
Ethnicity	2889.48	1	2889.48	102.22	0.26	***
Gender	576.17	1	576.17	20.38	0.06	***
Error	8395.60	297	28.27			

Note. Only significant effects are reported.

* = $p < 0.05$.

** = $p < 0.01$.

*** = $p < 0.001$.

In contrast to the treatment group findings, the control group used significantly *more* social-conventional justifications across the three time points ($F(2, 192) = 13.14, p < 0.001, \eta_p^2 = 0.12$; Fig. 3a), significantly *less* moral justifications from pre-test and post-test to 6-month follow-up ($F(2, 192) = 9.25, p < 0.001, \eta_p^2 = 0.09$; Fig. 3d), and significantly *less* excluder empathy justifications from pre-test to post-test—an effect which was maintained at follow-up ($F(2, 192) = 12.06, p < 0.001, \eta_p^2 = 0.11$; Fig. 3e). The control group's use of stereotype (Fig. 3b) and personal choice (Fig. 3c) justifications did not change significantly across the three time points. Cohen's *d* values comparing the differences between the treatment group and the control group averages from pre-test to post-test and pre-test to follow-up yielded moderate to large effect sizes for all significant findings. Effect sizes were particularly large for the skills group (see Table 3).

3.5. Secondary findings: gender and scenario effects

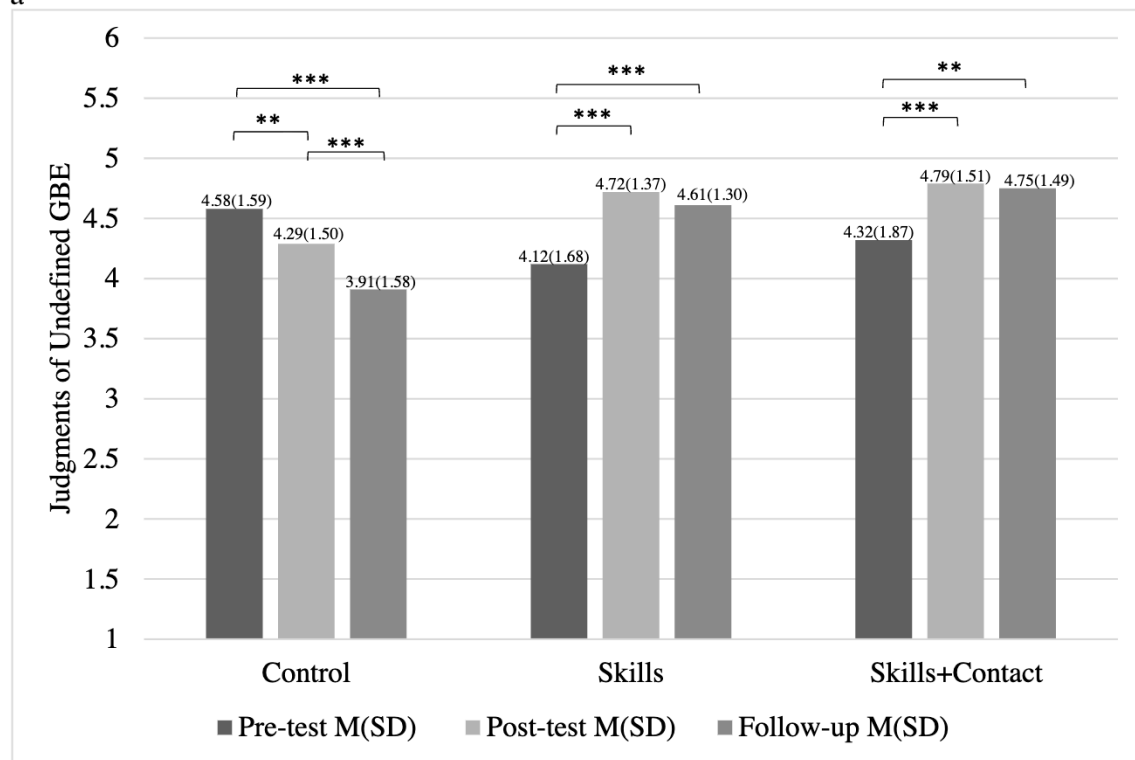
Though not directly related to intervention effectiveness, the following results help illuminate and/or replicate findings from the body of work on evaluations of intergroup exclusion, particularly in the Middle-East. Across the exclusion vignettes, girls were significantly more rejecting of exclusion than boys (see Tables 2, 4) and were more likely to use moral—and less likely to use social-conventional and stereotyped—justifications of their exclusion judgments than boys (online supplement Table 7a). For judgments, but not justifications, significant scenario effects emerged. In general for both undifferentiated and GBE, youth were more accepting of exclusion in the home context as compared to the peer context, an effect that was greater in magnitude for P–I participants (see Tables 2, 4). Finally, a scenario by justification interaction revealed that social-conventional justifications were used significantly more in the home than peer scenario ($F(1, 301) = 39.94, p < 0.001, \eta_p^2 = 0.12$), and moral justifications were used significantly more in the peer scenario than home scenario ($F(1, 301) = 23.03, p < 0.001, \eta_p^2 = 0.07$; online supplement Table 7a).

4. Discussion

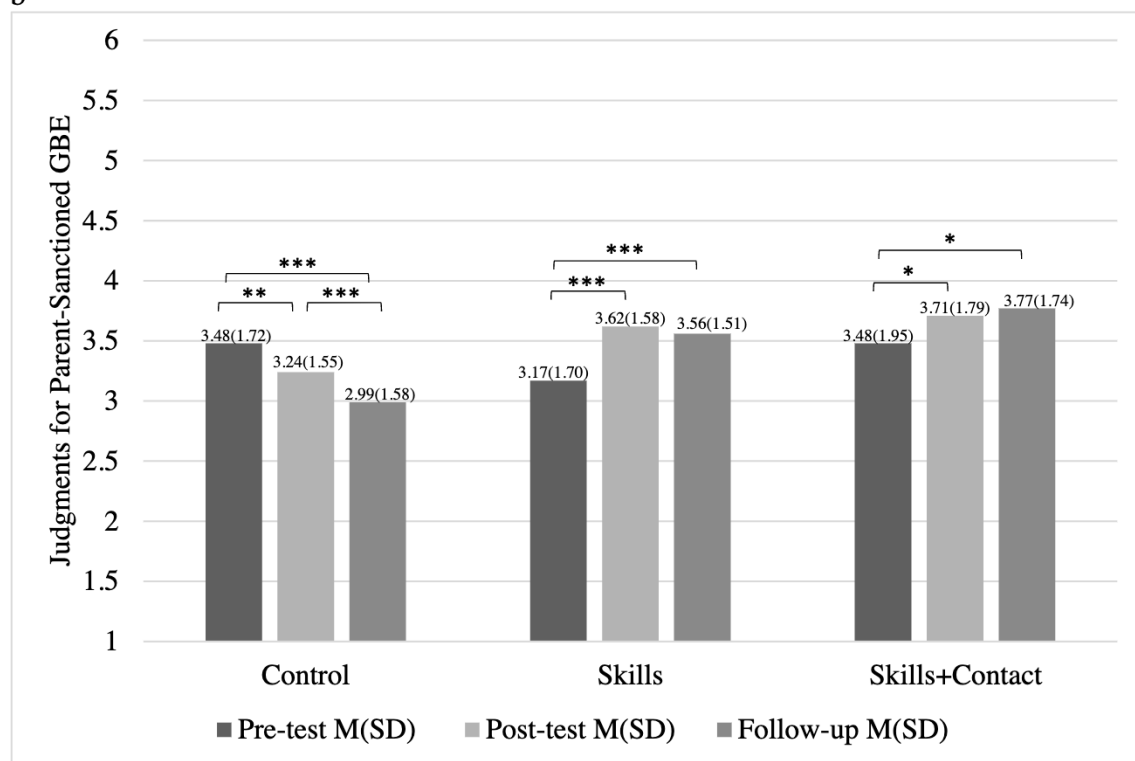
Several novel contributions to the field emerged from the current study. First and foremost, the findings overwhelmingly revealed that participants in both skills- and skills + contact-based intervention groups reduced prejudiced evaluations of intergroup social exclusion across time. At the same time, their counterparts in the control group showed increases in prejudice. Treatment group participants' prejudice reductions were reflected by their decreasing acceptance of outgroup exclusion, particularly when the exclusion was explicitly based on the outgroup member's ethnicity (both undefined and peer-sanctioned GBE). Additionally, over time, participants in both intervention groups were less likely to justify outgroup exclusion by appealing to stereotypes or social conventional justifications and were more likely to reject outgroup exclusion using empathic and moral justifications. That these effects were maintained at follow-up—even in the midst of an uptick in interethnic tension that has been named “the third Palestinian intifada” (Beaumont, 2015)—is demonstrative of the potential durability of both the skills and combined skills + contact interventions.

In stark contrast to the promising effects of the interventions, participants in the control group exhibited *increasing* acceptance of outgroup exclusion, whether undifferentiated or group-based. This suggests that without intervention, P–I and J–I youth in this area of conflict tend to grow increasingly accepting of prejudiced actions, both implicit and explicit, toward outgroup members. Likewise, participants in the control group appealed to social conventions to justify their ratings of exclusion significantly more often than all other reasons, and to empathic concerns and moral reasoning least frequently. This trend suggests that youth in this region recognize,

a



b



(caption on next page)

Fig. 2. Interaction Between Source, Time, and Group for GBE Judgments**a. Interaction Between Source, Time, and Group for Undefined GBE Judgments**

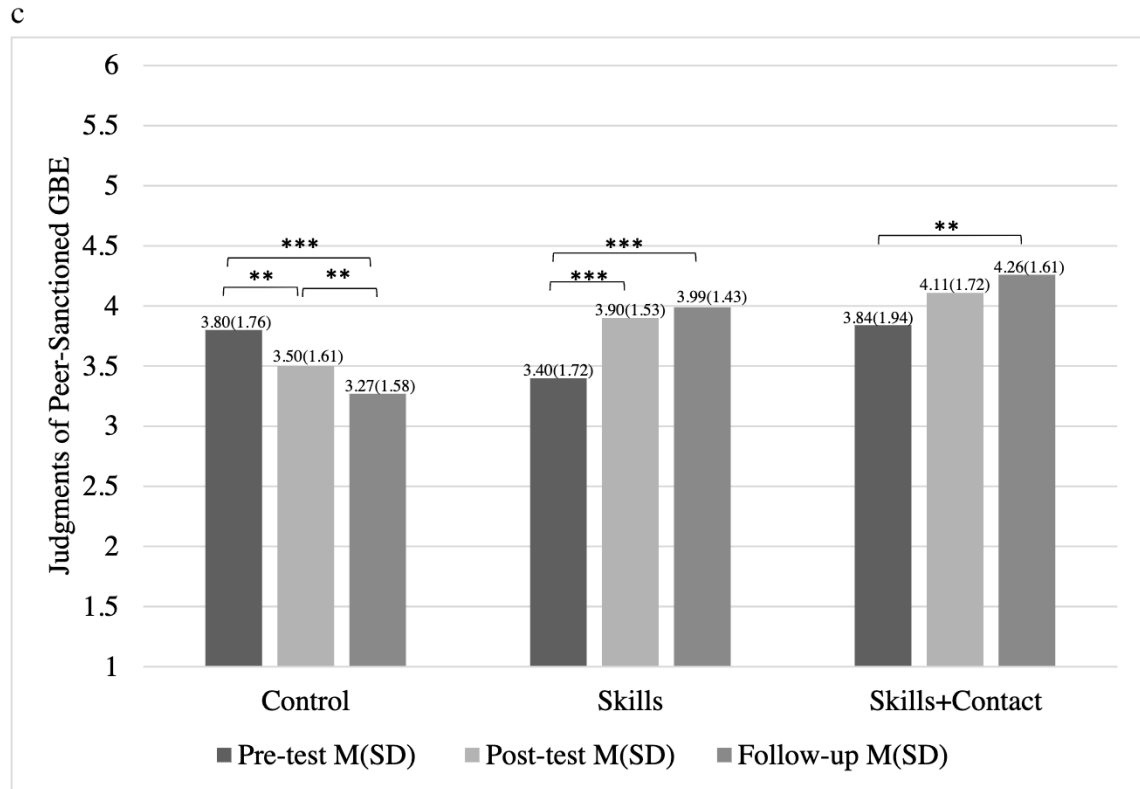
Note. Bars represent mean judgments of undefined group-based exclusion; Judgments were rated on a 7-point Likert-type scale (1, very, very good, to 7, very, very bad); Brackets indicate significant differences in judgments across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

b. Interaction Between Source, Time, and Group for Parent-Sanctioned GBE Judgments

Note. Bars represent mean judgments of parent-sanctioned group-based exclusion; Judgments were rated on a 7-point Likert-type scale (1, very, very good, to 7, very, very bad); Brackets indicate significant differences in judgments across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

c. Interaction Between Source, Time, and Group for Peer-Sanctioned GBE Judgments

Note. Bars represent mean judgments of peer-sanctioned group-based exclusion; Judgments were rated on a 7-point Likert-type scale (1, very, very good, to 7, very, very bad); Brackets indicate significant differences in judgments across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

**Fig. 2.** (continued)

and may be heavily influenced by, the prevailing social norms (see Brenick & Romano, 2016). Without intervention, they may perceive adherence to such norms as warranting outgroup exclusion (see Bar-Tal et al., 2017; Brenick & Romano, 2016).

Beyond simply demonstrating the effectiveness of both prejudice reduction interventions, the current study allowed for comparisons between the two approaches. Given the inconsistent or problematic findings of previous assessments of contact-based interventions in areas of conflict (see Barlow et al., 2012; Berger et al., 2015; Berger, Benatov, et al., 2016; Brenick et al., 2007; Guffler & Wagner, 2017; Maoz, 2011), and the promise of multi-theoretical or multi-component approaches (see Beelmann & Heinemann, 2014), the current study examined combining both skills and contact approaches in hopes that skills training would provide a needed supplement to contact to facilitate consistent and positive effects and to determine whether this combined approach was differentially effective as compared to skills training alone in an area of protracted conflict. Although largely similar in their effects, one noteworthy difference between the treatment groups emerged. Whereas participants in the skills group were less accepting of undifferentiated social exclusion when assessed immediately after the conclusion of the intervention (at post-test), it was not until the follow-up that participants in the skills + contact group exhibited such effects. Because participants in the skills + contact group were exposed to skills training and intergroup contact components simultaneously, it may be the case that it takes youth longer to apply the social-cognitive and social-emotional skills they learned in novel situations involving the outgroup. Future research could examine whether staggering skills and contact interventions might yield different effects.

Not only is this series of findings a significant contribution to the field, but so too are the methods through which they were obtained. The current study is one of only a handful of experimental, longitudinal assessments of in-school skills- and contact-based

Table 5
Results of repeated measures ANOVA for justifications.

	Sum of Squares	df	Mean Square	F	η_p^2	p
Significant within-subjects effects						
Justification	43.63	3	17.06	16.89	0.05	***
Justification X Group	40.14	5	7.85	7.77	0.05	***
Justification X Ethnicity	39.21	3	15.33	15.18	0.05	***
Justification X Gender	11.63	3	4.55	4.50	0.02	**
Error (Justification)	764.70	757	1.01			
Time	0.78	2	48	14.71	0.05	***
Time X Group	1.11	3	0.34	10.40	0.07	***
Error (Time)	15.72	489	0.03			
Justification X Scenario	2.43	4	0.70	2.69	0.01	*
Error (Justification X Scenario)	267.53	1037	0.26			
Time X Justification	8.81	4	2.38	9.38	0.03	***
Time X Justification X Group	27.46	16	3.71	14.63	0.09	***
Error (Time X Justification)	277.85	1095	0.25			
Significant between-subjects effects						
Group	5.51	2	2.75	21.11	0.13	***
Ethnicity	1.47	1	1.47	11.26	0.04	***
Error	38.56	296	0.13			

Note. Only significant effects are reported.

* = $p < 0.05$.

** = $p < 0.01$.

*** = $p < 0.001$.

prejudice reduction interventions implemented with groups outside of a laboratory setting and in an area of high conflict (for exceptions, see Berger et al., 2015; Berger, Benatov, et al., 2016). There have been several calls from the field for such research to be carried out beyond the walls of the laboratory, in real-world settings (see Beelmann & Heinemann, 2014; Paluck & Green, 2009) and particularly with groups in conflict (Kang, Gray, & Dovidio, 2014). Furthermore, this is the first study of its kind to assess the effectiveness of interventions in reducing prejudice as operationalized by evaluations of group-based social exclusion—a typical manifestation of prejudice in the lives of P–I and J–I youth (Bar-Tal et al., 2017; Brenick et al., 2007, 2010).

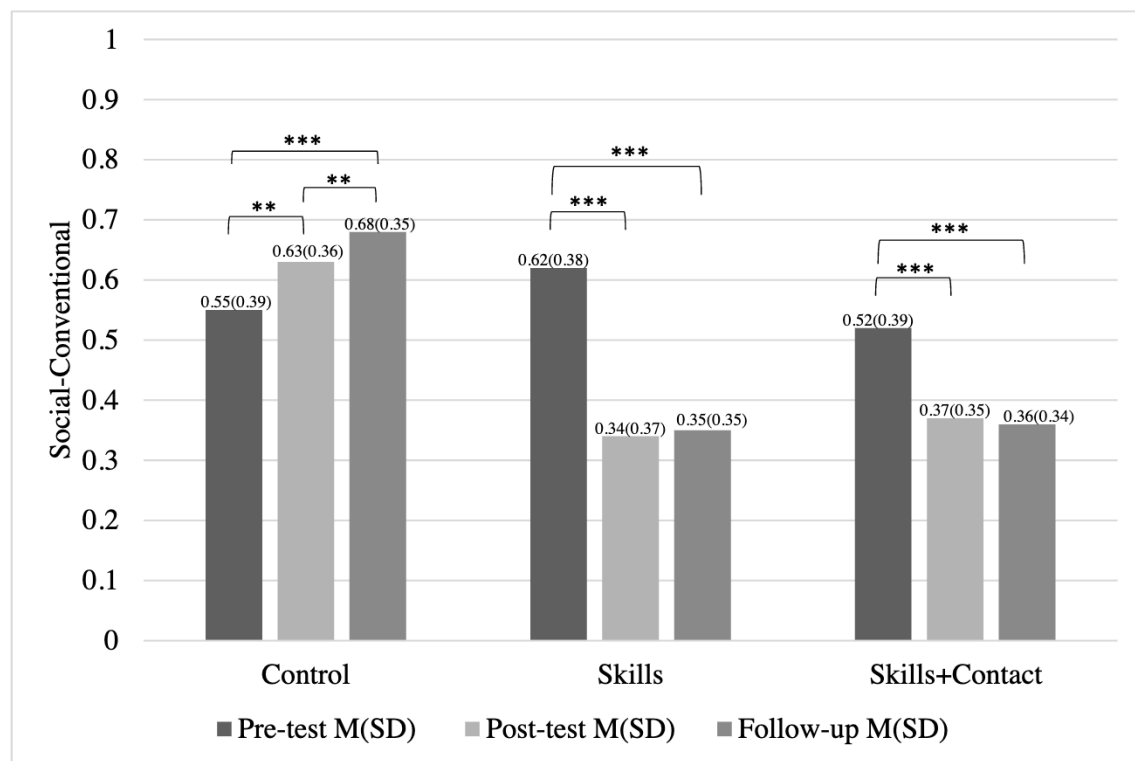
Finally, this methodological approach utilized multiple scenarios of intergroup social exclusion, and thus we were able to assess the effect of context—peer or home—on prejudiced attitudes. This distinction is critical in the Middle-East where cultural norms promote deference to parental authority (Alsamih & Tenenbaum, 2018). We found that, across time and treatment groups, P–I and J–I youth were more accepting of and more likely to appeal to social conventional reasons to justify outgroup exclusion from a sleepover in the family home than for exclusion from a peer activity outside of the home. These findings may be indicative of cultural norms and a developmental trend in which early adolescents still defer to their parents' social norms, particularly in the family domain while, in contrast, morality may take precedence in the peer domain (see Alsamih & Tenenbaum, 2018; Brenick & Romano, 2016).

Overall, both interventions are of merit and, depending on the environmental constraints of schools wishing to implement such programs, either treatment method may be appropriate. For instance, if outgroup contact is infeasible, the skills treatment would be apropos. Conversely, if schools or communities are working toward desegregation, supplementing intergroup contact with other approaches to prejudice reduction, such as skills-based interventions, may help reduce the potential for intergroup contact to be negative or inconsistent in its benefits, as observed in previous studies examining the impacts of contact alone (e.g., Brown & Hewstone, 2005; Salomon, 2006; Vorauer & Sasaki, 2011). The findings also point to the importance of context when evaluating the effectiveness of interventions with youth; we must further assess how peer-based intergroup contact generalizes to contexts governed by adults in the family or community.

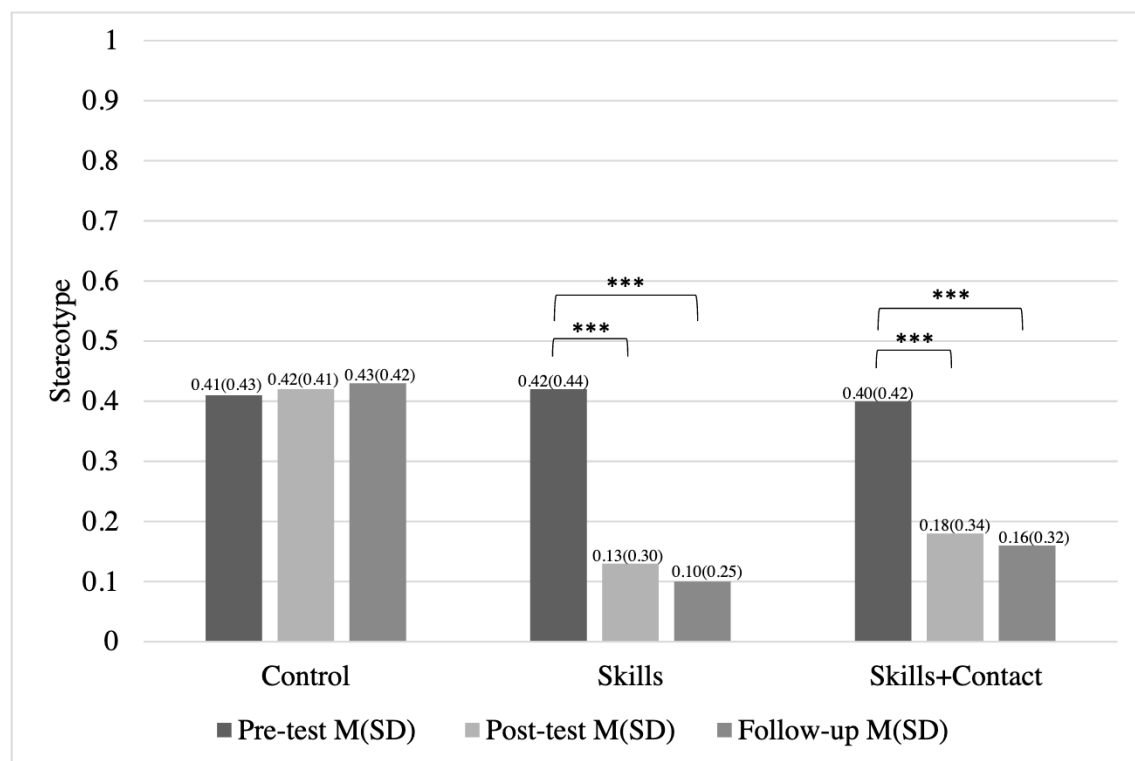
4.1. Limitations of the study and areas for future research

Although promising, this study's findings should be considered in light of the following shortcomings. First, because the current investigation was carried out in a single city in a high conflict region, the findings may not generalize to other settings. Future research should replicate this study in various cities that may be more or less affected by this or other ongoing conflicts. The study could also be replicated in less conflict-laden settings. Second, the study's findings were based on self-report data collected in a survey format. Although the instruments utilized herein have good psychometric properties, self-report measures can be subject to bias. Further, this survey alone may not have been sensitive enough to detect differences between the intervention groups and could not

a



b



(caption on next page)

Fig. 3. Interaction Between Time, Group and Justification Type**a. Average Use of Social-Conventional Justification Across Time and Treatment Group**

Note. Bars represent mean proportion use of justification; Brackets indicate significant differences in use of social-conventional justifications across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

b. Average Use of Stereotype Justification Across Time and Treatment Group

Note. Bars represent mean proportion use of justification; Brackets indicate significant differences in use of stereotype justifications across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

c. Average Use of Moral Justifications Across Time and Treatment Group

Note. Bars represent mean proportion use of justification; Brackets indicate significant differences in use of moral justifications across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

d. Average Use of Excluder Empathy Justifications Across Time and Treatment Group

Note. Bars represent mean proportion use of justification; Brackets indicate significant differences in use of excluder empathy justifications across time points; * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

3e. Average Use of Personal Choice Justifications Across Time and Treatment Group

Note. Bars represent mean proportion use of justification; No significant differences observed across time and treatment groups in use of personal choice justifications.

determine whether the attitudinal changes manifested by participants led to changes in actual behavior. Thus, future studies could supplement the self-administered questionnaires with behavioral observations, to provide multiple sources of data and determine if behavioral changes accompanied changes in attitudes.

Third, baseline assessments were not equal for P–I and J–I samples, thus requiring that we control for ethnicity rather than assess it as a variable of interest. Future studies should attempt to find groups equivalent at baseline to allow for such comparisons. Fourth, though we attempted to create equal status for P–I and J–I participants within the contact activities of the skills + contact intervention, participants' perceptions of their comparative status were not assessed. Because of this, we cannot be certain that the inequality experienced by P–I outside of the contact setting was not brought with them into this setting.

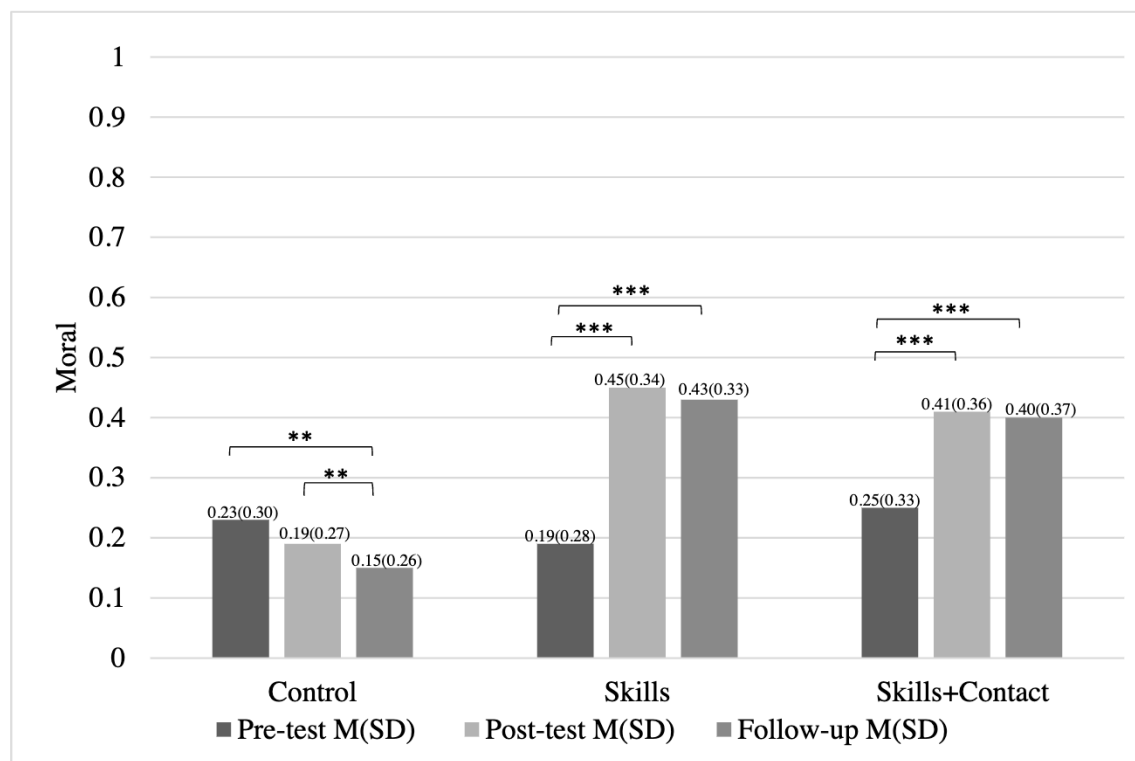
Fifth, our data were multilevel in nature with the interventions administered at the classroom level. Although the calculated ICC values and wild cluster bootstrapping analyses allow for non-multilevel analyses, future research should utilize such methods if clustering effects are significant and, as a result, will benefit from directly assessing the impact of classroom differences, particularly with the few instances in which the confidence intervals increased with wild cluster bootstrapping. Further inspection of the ICCs revealed higher clustering effects in the home compared to the peer context, and in judgments of exclusion compared to the justifications. Future studies evaluating the impact of prejudice interventions across classrooms should strive to sample more classrooms as a further precaution for reducing this clustering effect in their sample. In the present study, there was a limited number of classrooms in a limited number of schools, which prevented the authors from increasing the number of clusters—a matter that may not be an issue in other regions.

Finally, the interventions were designed to be easily and inexpensively incorporated into a school curriculum. However, our interventions were administered by trained research assistants who were brought into the schools. For these programs to be scalable and cost-effective for implementation in schools, it is critical to train teachers to the point of feeling confident in delivering the intervention content and integrating it into their current curricula.

4.2. Implications of the study

This study's findings lend support to the short-term effectiveness of both skills- and combined skills- and contact-based approaches in reducing prejudice among children in an area of conflict. Further, the findings suggest that the benefits of both prejudice reduction interventions persist even after the conclusion of the interventions. In some contexts, contact interventions may be infeasible—or even incendiary if the contact is negative (Brown & Hewstone, 2005; Salomon, 2006; Vorauer & Sasaki, 2011). In the current study, however, we see that contact supplemented with a skills training component largely reduces prejudice and promotes positive intergroup attitudes, offering even those students in segregated schools the opportunity to engage in positive contact with the outgroup. However, in cases in which contact cannot or will not be established, our findings demonstrate that the skills intervention alone can also yield positive intergroup effects and may be a preferred alternative. Thus, schools aiming to reduce outgroup prejudice and promote positive intergroup attitudes would do well to implement one of these intervention programs, depending on their resources and context. In fact, because learning from a role model is a key factor in promoting positive intergroup attitudes, training teachers—figures with whom students often have close relationships and see as role models—to administer either of these interventions within a school setting might further facilitate positive outcomes. Schools—institutions in which youth interact daily—taking an active role in the promotion of positive intergroup attitudes and the reduction of prejudice might validate and inspire opposition to the inequality and prejudice maintained by the status quo.

c



d

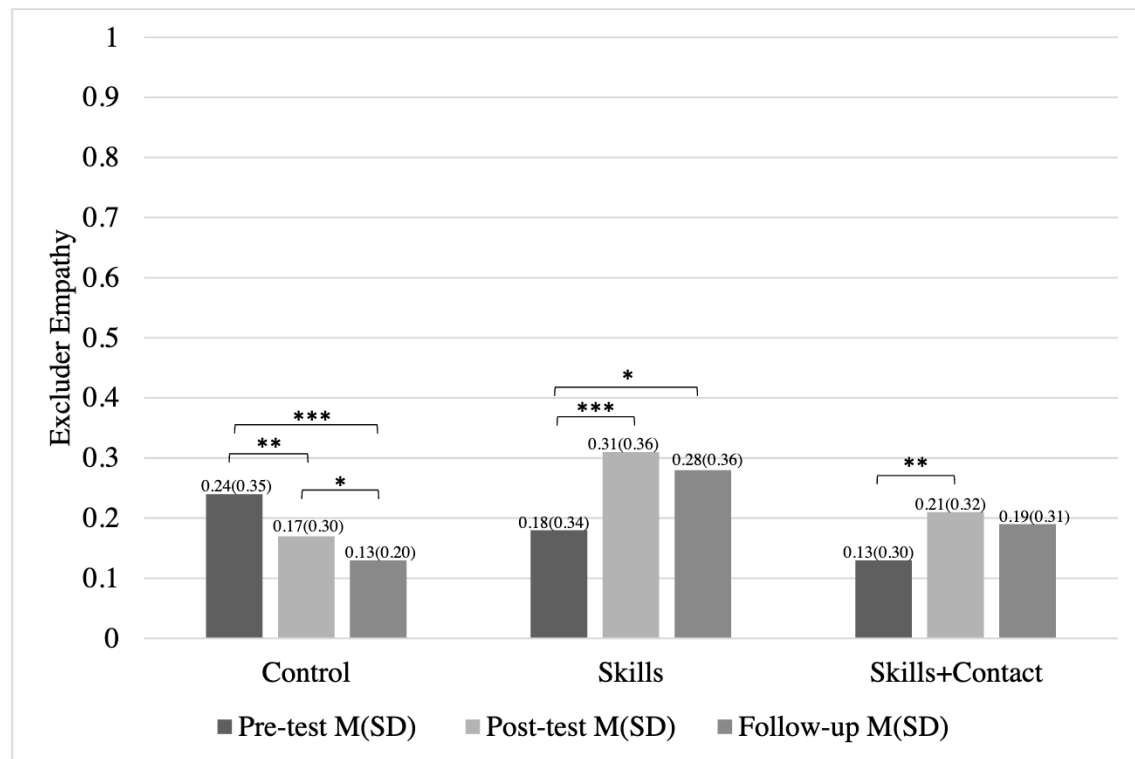


Fig. 3. (continued)

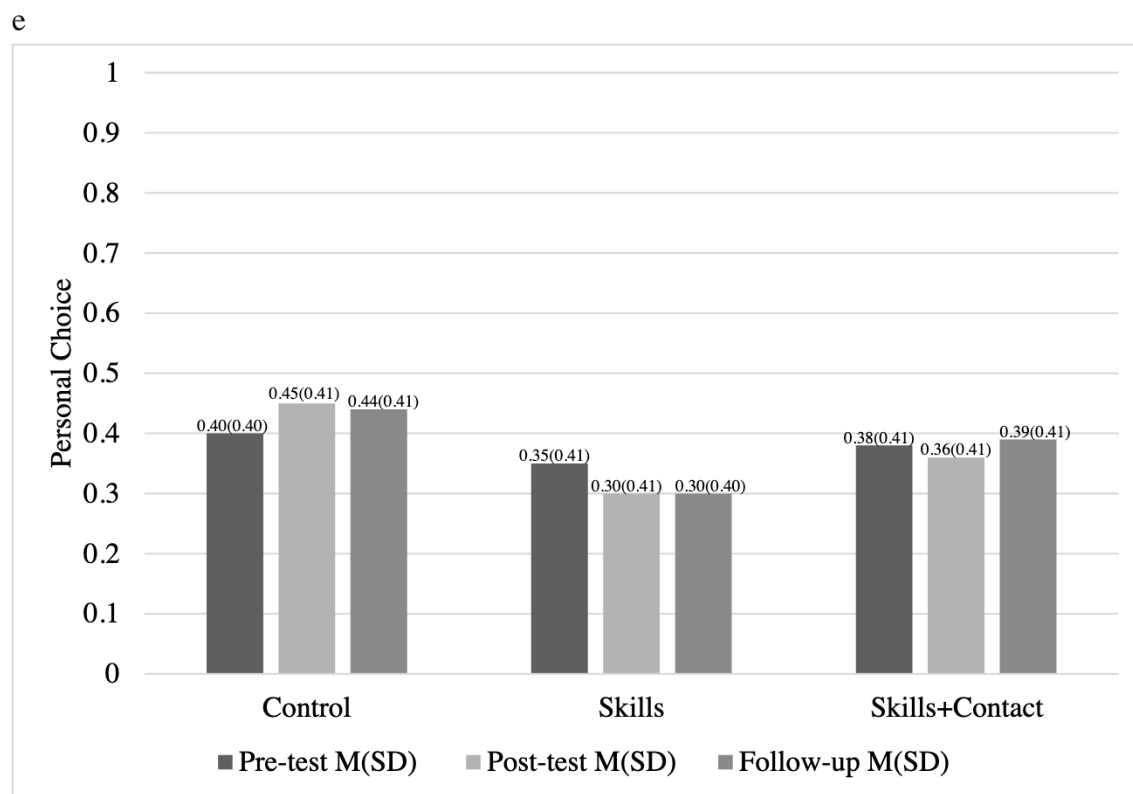


Fig. 3. (continued)

Acknowledgement

Partial funding for the project was provided by the University of Connecticut (UConn) Scholarship Facilitation Fund Grant (USA) and the Cohen Family UConn-Israel Research Start-up Grant (USA) awarded to the first author.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jsp.2019.07.001>.

References

- Aboud, F. E. (2008). A social-cognitive developmental theory of prejudice. In S. M. Quintana, C. McKown, S. M. Quintana, & C. McKown (Eds.). *Handbook of race, racism, and the developing child* (pp. 55–71). Hoboken, NJ, US: John Wiley & Sons Inc.
- 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, US: Lawrence Erlbaum Associates Publishers.
- Aboud, F. E., Tredoux, C., Tropp, L. R., Brown, C. S., Niens, U., & Noor, N. M. (2012). Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review. *Developmental Review*, 32, 307–336. <https://doi.org/10.1016/j.dr.2012.05.001>.
- Abrams, D., Palmer, S. B., Rutland, A., Cameron, L., & Van de Vyver, J. (2014). Evaluations of and reasoning about normative and deviant ingroup and outgroup members: Development of the black sheep effect. *Developmental Psychology*, 50, 258–270. <https://doi.org/10.1037/a0032461>.
- Allport, G. (1954). *The nature of prejudice*. Oxford, England: Addison-Wesley.
- Alsamir, M., & Tenenbaum, H. R. (2018). Saudi Arabian children's reasoning about religion-based exclusion. *British Journal of Developmental Psychology*. <https://doi.org/10.1111/bjdp.12238> Advance online publication.
- Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom: Building cooperation in the classroom* (2nd ed.). New York, NY: Addison Wesley Longman.
- Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. M., Harwood, J., ... Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, 38, 1629–1643. <https://doi.org/10.1177/0146167212457953>.
- Bar-Tal, D. (2011). *Intergroup conflicts and their resolution: Social psychological perspective*. New York, NY: Psychology Press.
- Bar-Tal, D. (2013). *Intractable conflicts: Socio-psychological foundations and dynamics*. Cambridge, UK: Cambridge University Press <https://doi.org/10.1111/pops.12358>.
- Bar-Tal, D., Diamond, A. H., & Nasie, M. (2017). Political socialization of young children in intractable conflicts: Conception and evidence. *International Journal of Behavioral Development*, 41, 415–425. <https://doi.org/10.1177/0165025416652508>.
- Bar-Tal, D., & Teichman, Y. (2005). *Stereotypes and prejudice in conflict: Representations of Arabs in Israeli Jewish society*. New York, NY: Cambridge University Press <https://doi.org/10.1017/CBO9780511499814>.
- Beaumont, P. (2015). Is the third Palestinian intifada on the way - or has it already begun? The Guardian. October 5. Retrieved from <https://www.theguardian.com/world/2015/oct/05/third-palestinian-intifada-on-way-or-already-begun>.
- Beelmann, A., & Heinemann, K. S. (2014). Preventing prejudice and improving intergroup attitudes: A meta-analysis of child and adolescent training programs. *Journal*

- of *Applied Developmental Psychology*, 35, 10–24. <https://doi.org/10.1016/j.appdev.2014.06.005>.
- Bekerman, Z. (2005). Complex contexts and ideologies: Bilingual education in conflict-ridden areas. *Journal of Language, Identity, and Education*, 4, 1–20. https://doi.org/10.1207/s15327701jlie0401_1.
- Benner, A. D., Crosnoe, R., & Eccles, J. S. (2015). Schools, peers, and prejudice in adolescence. *Journal of Research on Adolescence*, 25, 173–188. <https://doi.org/10.1111/jora.12106>.
- Berger, R., Abu-Raiya, H., & Gelkopf, M. (2015). The art of living together: Reducing prejudicial attitudes through the Arab–Jewish class exchange program (CEP). *Journal of Educational Psychology*, 107, 678–688. <https://doi.org/10.1037/edu0000015>.
- Berger, R., Benatov, J., Abu-Raiya, H., & Tadmor, C. T. (2016). Reducing prejudice and promoting positive intergroup attitudes among elementary-school children in the context of the Israeli–Palestinian conflict. *Journal of School Psychology*, 57, 53–72. <https://doi.org/10.1016/j.jsp.2016.04.003>.
- Berger, R., Brenick, A., & Tarrasch, R. (2018). Reducing stereotyping and prejudicial attitudes in Israeli-Jewish elementary school children with a mindfulness and compassion-based social-emotional program. *Mindfulness*, 9, 1768–1779. <https://doi.org/10.1007/s12671-018-0919-y>.
- Berger, R., Gelkopf, M., Heineberg, Y., & Zimbardo, P. (2016). A school-based intervention for reducing posttraumatic symptomatology and intolerance during political violence. *Journal of Educational Psychology*, 108, 761–771. <https://doi.org/10.1037/edu0000066>.
- Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, 16, 162–166.
- Brenick, A., & Killen, M. (2014). Moral judgments about Jewish-Arab intergroup exclusion: The role of culture and contact. *Developmental Psychology*, 50, 86–99.
- Brenick, A., Killen, M., Lee-Kim, J., Fox, N., Leavitt, L., Raviv, A., & Al-Smadi, Y. (2010). Social understanding in young Israeli-Jewish, Israeli-Palestinian, Palestinian, and Jordanian children: Moral judgments and stereotypes. *Early Education and Development*, 21, 886–911. <https://doi.org/10.1080/10409280903236598>.
- Brenick, A., Lee-Kim, J., Killen, M., Fox, N. A., Raviv, A., & Leavitt, L. A. (2007). Social judgments in Israeli and Arab children: Findings from media-based intervention projects. In D. Lemish, & M. Götz (Eds.). *Children and media at times of conflict and war* (pp. 287–308). Cresskill, NJ: Hampton Press.
- Brenick, A., & Romano, K. (2016). Perceived peer and parent outgroup norms, cultural identity, and adolescents' reasoning about peer intergroup exclusion. *Child Development*, 87, 1392–1408.
- Brenick, A., & Halgunseth, L. (2017). Applying developmental intergroup perspectives to the social ecologies of bullying: Lessons from developmental social psychology. *Journal of Adolescence*, 59, 90–95. <https://doi.org/10.1016/j.adolescence.2017.05.018>.
- Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in Experimental Social Psychology*, 37, 255–343. [https://doi.org/10.1016/S0065-2601\(05\)37005-5](https://doi.org/10.1016/S0065-2601(05)37005-5).
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90, 414–427.
- Cameron, L., & Rutland, A. (2006). Extended contact through story reading in school: Reducing children's prejudice toward the disabled. *Journal of Social Issues*, 62, 469–488. <https://doi.org/10.1111/j.1540-4560.2006.00469.x>.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.
- Cooley, S., Elenbaas, L., & Killen, M. (2016). Social exclusion based on group membership is a form of prejudice. In S. S. Horn, M. D. Ruck, & L. S. Liben (Eds.). *Equity and justice in developmental science: Implications for young people, families, and communities* (pp. 103–129). San Diego, CA, US: Elsevier Academic Press.
- 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. <https://doi.org/10.1177/108886831141110>.
- Dovidio, J. F., Gaertner, S. L., & Saguy, T. (2007). Another view of “we”: Majority and minority group perspectives on a common ingroup identity. *European Review of Social Psychology*, 18, 296–330.
- Doyle, A. B., & Aboud, F. E. (1995). A longitudinal study of white children's racial prejudice as a social-cognitive development. *Merrill-Palmer Quarterly*, 41, 209–228.
- Fischer, R., & Fontaine, J. R. J. (2011). Methods for investigating structural equivalence. In D. Matsumoto, & F. J. R. van de Vijver (Eds.). *Culture and psychology. Cross-cultural research methods in psychology* (pp. 179–215). New York, NY, US: Cambridge University Press.
- Frey, K. S., Nolen, S. B., Van Schoiack Edstrom, L., & Hirschstein, M. K. (2005). Effects of a school-based social-emotional competence program: Linking children's goals, attributions, and behavior. *Journal of Applied Developmental Psychology*, 26, 171–200.
- Graham, N., Mahmood, A., & Björn, H. (2016). *Multiwaycov: Multi-way standard error clustering*. (R-package).
- Guffler, K., & Wagner, U. (2017). Backfire of good intentions: Unexpected long-term contact intervention effects in an intractable conflict area. *Peace and Conflict: Journal of Peace Psychology*, 23, 383–391. <https://doi.org/10.1037/pac0000264>.
- Horn, S. S., & Nucci, L. (2006). Harassment of gay and lesbian youth and school violence in America: An analysis and directions for intervention. In C. Daiute, Z. Beykont, C. Higson-Smith, & L. Nucci (Eds.). *International perspectives on youth conflict and development* (pp. 139–155). New York, NY, US: Oxford University Press.
- Jugert, P., & Peddes, A. R. (2017). Children's and adolescents' cross-ethnic friendships. In A. Rutland, D. Nesdale, & C. S. Brown (Eds.). *The Wiley-Blackwell handbook of group processes in children and adolescents* (pp. 373–392). Chichester: Wiley-Blackwell.
- Kang, Y., Gray, J. R., & Dovidio, J. F. (2014). The nondiscriminating heart: Lovingkindness meditation training decreases implicit intergroup bias. *Journal of Experimental Psychology: General*, 143, 1306–1313. <https://doi.org/10.1037/a0034150>.
- Killen, M., & Rutland, A. (2011). *Children and social exclusion: Morality, prejudice, and group identity*. Hoboken, NJ: Wiley-Blackwell <https://doi.org/10.1002/9781444396317>.
- Lemmer, G., & Wagner, U. (2015). Can we really reduce ethnic prejudice outside the lab? A meta-analysis of direct and indirect contact interventions. *European Journal of Social Psychology*, 45, 152–168. <https://doi.org/10.1002/ejsp.2079>.
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2, 57–64.
- Malti, T., Killen, M., & Gasser, L. (2012). Social judgments and emotion attributions about exclusion in Switzerland. *Child Development*, 83, 697–711.
- Malti, T., & Noam, G. G. (2016). Social-emotional development: From theory to practice. *European Journal of Developmental Psychology*, 13, 652–665. <https://doi.org/10.1080/17405629.2016.1196178>.
- Maaz, I. (2011). Contact in protracted asymmetrical conflict: Twenty years of planned encounters between Israeli Jews and Palestinians. *Journal of Peace Research*, 48, 115–152.
- Morris, S. B. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11, 364–386. <https://doi.org/10.1177/1094428106291059>.
- Nesdale, D. (2004). Social identity processes and children's ethnic prejudice. In M. Bennett, & F. Sani (Eds.). *The development of the social self* (pp. 219–246). New York, NY: Psychology Press.
- Nucci, L., & Turiel, E. (2009). Capturing the complexity of moral development and education. *Mind, Brain, and Education*, 3, 151–159. <https://doi.org/10.1111/j.1751-228X.2009.01065.x>.
- Palmer, S. B., & Abbott, N. (2017). Bystander responses to bias-based bullying in schools: A developmental intergroup approach. *Child Development Perspectives*, 12, 39–44.
- 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. <https://doi.org/10.1146/annurev.psych.60.110707.163607>.
- Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49, 65–85.
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90, 751–783. <https://doi.org/10.1037/0022-3514.90.5.751>.
- 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. <https://doi.org/10.1111/j.1467-8624.2011.01668.x>.
- Ruck, M. D., Park, H., Crystal, D. S., & Killen, M. (2015). Intergroup contact is related to evaluations of interracial peer exclusion in African American students. *Journal of Youth and Adolescence*, 44, 1226–1240. <https://doi.org/10.1007/s10964-014-0227-3>.

- Rutland, A., & Killen, M. (2015). A developmental science approach to reducing prejudice and social exclusion: Intergroup processes, social-cognitive development, and moral reasoning. *Social Issues and Policy Review*, 9, 121–154. <https://doi.org/10.1111/sipr.12012>.
- Rutland, A., Killen, M., & Abrams, D. (2010). A new social-cognitive developmental perspective on prejudice: The interplay between morality and group identity. *Perspectives on Psychological Science*, 5, 279–291. <https://doi.org/10.1177/1745691610369468>.
- Salomon, G. (2006). Does peace education really make a difference? *Peace and Conflict: Journal of Peace Psychology*, 12, 37–48. https://doi.org/10.1207/s15327949pac1201_3.
- Save the Children (2006). *The state of the world's children: Excluded and invisible*. New York, NY: UNICEF. Retrieved from https://www.unicef.org/sowc06/pdfs/sowc06_fullreport.pdf.
- Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychological Bulletin*, 140, 921–948. <https://doi.org/10.1037/a0035754>.
- Shor, N. (2017). Don't be fooled by its Israeli Jewish gentrification - Jaffa largely remains an occupied military zone. Israel: Haretz. August 10. Retrieved from <https://www.haaretz.com/israel-news/premium-don-t-be-fooled-by-jaffa-s-israeli-jewish-gentrification-1.5441838>.
- Smetana, J. G. (2006). Social-cognitive domain theory: Consistencies and variations in children's moral and social judgments. In M. Killen, & J. G. Smetana (Eds.). *Handbook of moral development* (pp. 119–153). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Smetana, J. G., Jambon, M., & Ball, C. (2014). The social domain approach to children's moral and social judgments. In M. Killen, & J. G. Smetana (Eds.). *Handbook of moral development* (pp. 23–45). (2nd ed.). New York, NY, US: Psychology Press.
- Taylor, S. E. (2015). Social cognition and health. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Vol. Eds.), *Attitudes and social cognition*: . Vol. 1. *APA handbooks in psychology. APA handbook of personality and social psychology* (pp. 339–361). Washington, DC: American Psychological Association.
- Tenenbaum, H. R., Leman, P., & Aznar, A. (2017). Children's reasoning about peer and school segregation in a diverse society. *Journal of Community & Applied Social Psychology*, 27, 358–365. <https://doi.org/10.1002/casp.2311>.
- Titzmann, P. F., Brenick, A., & Silbereisen, R. K. (2015). Friendships fighting prejudice: A longitudinal perspective on adolescents' cross-group friendships with immigrants. *Journal of Youth and Adolescence*, 44, 1318–1331. <https://doi.org/10.1007/s1096401502566>.
- Tobler, A. L., Maldonado-Molina, M. M., Staras, S. A. S., O'Mara, R. J., Livingston, M. D., & Komro, K. A. (2013). Perceived racial/ethnic discrimination, problem behaviors, and mental health among minority urban youth. *Ethnicity & Health*, 18, 337–349. <https://doi.org/10.1080/13557858.2012.730609>.
- Tropp, L. R., & Prenovost, M. A. (2008). The role of intergroup contact in predicting children's interethnic attitudes: Evidence from meta-analytic and field studies. In S. R. Levy, & M. Killen (Eds.). *Intergroup attitudes and relations in childhood through adulthood* (pp. 236–248). New York, NY: Oxford University Press.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, England: Cambridge University Press.
- Vorauer, J. D., & Sasaki, S. J. (2011). In the worst rather than the best of times: Effects of salient intergroup ideology in threatening intergroup interactions. *Journal of Personality and Social Psychology*, 101, 307–320. <https://doi.org/10.1037/a0023152>.

Table 1a

Between- and Within-cluster variance, and Intraclass Correlations for Judgments and Justifications in the Peer and Home Contexts

Outcomes	S_b^2	S_w^2	ρ	ESS
<u>Peer Context Judgments</u>				
Undifferentiated Exclusion	0.37	2.27	0.14	303
Group-based Exclusion (GBE)	0.49	2.57	0.16	304
Parent-sanctioned GBE	0.73	3.28	0.18	304
Peer-sanctioned GBE	0.62	3.25	0.16	304
<u>Peer Context Justifications</u>				
Societal	0.002	0.18	0.01	300
Victim Blaming	0.02	0.22	0.07	302
Personal Choice	0.0003	0.24	0.001	300
Moral	0.01	0.11	0.07	302
Excluder Empathy	0.01	0.15	0.05	301
<u>Home Context Judgments</u>				
Undifferentiated Exclusion	0.73	2.53	0.22	305
Group-based Exclusion	0.78	2.89	0.21	305
Parent-sanctioned GBE	0.82	2.71	0.23	306
Peer-sanctioned GBE	0.73	2.90	0.20	305
<u>Home Context Justifications</u>				
Societal	0.01	0.18	0.04	301
Victim Blaming	0.01	0.24	0.02	301
Personal Choice	1.437×10^{-15}	0.23	0.00	300
Moral	0.01	0.10	0.05	301
Excluder Empathy	0.73	2.90	0.20	305

Note. S_b^2 = between-cluster variance, S_w^2 = within-cluster variance, ρ = intraclass correlation; ESS = effective sample size. Variance components were calculated using the lmer function of the lme4 package in r. Intraclass correlations were calculated using the formula $\rho = S_b^2 / (S_b^2 + S_w^2)$ (see Killip, Mahfoud, & Pearce, 2004).

Table 2a

Wild Cluster Bootstrapping for Intervention Effects on Undifferentiated Exclusion Judgments

Group	Pre-test		Post-test		6-month follow-up	
	Pre-Bootstrap	Wild Cluster Bootstrap	Pre-Bootstrap	Wild Cluster Bootstrap	Pre-Bootstrap	Wild Cluster Bootstrap
	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI
<u>Peer Context</u>						
Skills	-.47 (.21)* [-.88, -.06]	-.47 (.12)*** [-.71, -.23]	-.14 (.19) [-.51, .23]	-.14 (.11) [-.36, .08]	.11 (.18) [-.25, .47]	.11 (.14) [-.17, .39]
Skills + Contact	-.07 (.21) [-.48, .34]	-.07 (.23) [-.52, .39]	.23 (.19) [-.14, .60]	.23 (.20) [-.16, .62]	.50 (.18)** [.14, .86]	.50 (.16)** [.19, .82]
<u>Home Context</u>						
Skills	-.43 (.22)+ [-.87, .01]	-.43 (.22)+ [-.87, .01]	.17 (.20) [-.24, .57]	.17 (.20) [-.24, .57]	.35 (.19)+ [-.02, .72]	.35 (.17)* [.02, .69]
Skills + Contact	.11 (.22) [-.34, .55]	.11 (.26) [-.40, .62]	.63 (.21)** [.23, 1.04]	.63 (.26)* [.12, 1.14]	.90 (.19)** [.53, 1.27]	.90 (.19)** [.53, 1.27]

Note. Regression models for treatment group effects before and after wild cluster bootstrapping (Cameron, Gelbach, & Miller, 2008), at pre-test, post-test, and 6-month follow-up. Ethnicity and gender were controlled for in the models. Referent group = control group. All confidence intervals are calculated as 95% confidence intervals. *B* = treatment effect slope, *SE* = standard error, CI = confidence interval.

+*p* < .10 **p* < .05. ***p* < .01. ****p* < .001.

Table 3a

Wild Cluster Bootstrapping for Intervention Effects on Undefined Group-based Exclusion Judgments

Group	Pre-test		Post-test		6-month follow-up	
	Pre-Bootstrap	Wild Cluster Bootstrap	Pre-Bootstrap	Wild Cluster Bootstrap	Pre-Bootstrap	Wild Cluster Bootstrap
	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI
<u>Peer Context</u>						
Skills	-.50 (.22)* [-.94, -.06]	-.50 (.23)* [-.95, -.05]	.40 (.18)* [.04, .76]	.40 (.16)* [.09, .72]	.74 (.19)*** [.37, 1.10]	.74 (.15)*** [.45, 1.04]
Skills + Contact	-.36 (.22) [-.80, .08]	-.36 (.35) [-1.05, .33]	.41 (.18)* [.05, .77]	.41 (.20)* [.02, .80]	.76 (.19)*** [.40, 1.13]	.76 (.19)*** [.40, 1.13]
<u>Home Context</u>						
Skills	-.47 (.24)* [-.94, -.001]	-.47 (.20)* [-.86, -.08]	.43 (.21)* [.01, .84]	.43 (.15)** [.14, .73]	.64 (.20)** [.24, 1.04]	.64 (.14)*** [.37, .92]
Skills + Contact	-.12 (.24) [-.59, .35]	-.12 (.27) [-.65, .41]	.63 (.21)** [.22, 1.04]	.63 (.20)** [.24, 1.02]	.97 (.20)*** [.57, 1.37]	.97 (.17)*** [.64, 1.31]

Note. Regression models for treatment group effects before and after wild cluster bootstrapping (Cameron, Gelbach, & Miller, 2008), at pre-test, post-test, and 6-month follow-up. Ethnicity and gender were controlled for in the models. Referent group = control group. All confidence intervals are calculated as 95% confidence intervals. *B* = treatment effect slope, *SE* = standard error, CI = confidence interval.

+*p* < .10 **p* < .05. ***p* < .01. ****p* < .001.

Table 4a

Wild Cluster Bootstrapping for Intervention Effects on Peer-sanctioned Group-based Exclusion Judgments

Group	Pre-test		Post-test		6-month follow-up	
	Pre-Bootstrap	Wild Cluster Bootstrap	Pre-Bootstrap	Wild Cluster Bootstrap	Pre-Bootstrap	Wild Cluster Bootstrap
	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI
<u>Peer Context</u>						
Skills	-.49 (.25)* [-.98, -.01]	-.49 (.14)*** [-.77, -.22]	.39 (.22) ⁺ [-.04, .82]	.39 (.17)* [.06, .73]	.81 (.21)*** [.40, 1.22]	.81 (.17)*** [.48, 1.15]
Skills + Contact	-.01 (.25) [-.49, .48]	-.01 (.17) [-.35, .33]	.61 (.22)** [.18, 1.04]	.61 (.20)** [.22, 1]	.99 (.21)*** [.58, 1.40]	.99 (.13)*** [.73, 1.25]
<u>Home Context</u>						
Skills	-.34 (.24) [-.81, .13]	-.34 (.19) ⁺ [-.71, .03]	.37 (.22) ⁺ [-.06, .79]	.37 (.17)* [.04, .71]	.60 (.20)** [.20, .99]	.60 (.16)*** [.29, .92]
Skills + Contact	.15 (.24) [-.32, .62]	.15 (.19) [-.22, .52]	.65 (.22)** [.23, 1.08]	.65 (.21)** [.24, 1.06]	1.06 (.20)*** [.66, 1.45]	1.06 (.17)*** [.73, 1.40]

Note. Regression models for treatment group effects before and after wild cluster bootstrapping (Cameron, Gelbach, & Miller, 2008), at pre-test, post-test, and 6-month follow-up. Ethnicity and gender were controlled for in the models. Referent group = control group. All confidence intervals are calculated as 95% confidence intervals. *B* = treatment effect slope, *SE* = standard error, CI = confidence interval.

⁺p < .10 *p < .05. **p < .01. ***p < .001.

Table 5a

Wild Cluster Bootstrapping for Intervention Effects on Parent-sanctioned Group-based Exclusion Judgments

Group	Pre-test		Post-test		6-month follow-up	
	Pre- Bootstrap	Wild Cluster Bootstrap	Pre- Bootstrap	Wild Cluster Bootstrap	Pre- Bootstrap	Wild Cluster Bootstrap
	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI	<i>B (SE)</i> CI
<u>Peer Context</u>						
Skills	-.47 (.25) ⁺ [-.97, .03]	-.47 (.20)* [-.86, -.08]	.28 (.23) [-.18, .73]	.28 (.17) ⁺ [-.06, .62]	.63 (.22)** [.19, 1.07]	.63 (.16)** [.32, .95]
Skills + Contact	-.09 (.26) [-.59, .41]	-.09 (.28) [-.64, .46]	.32 (.23) [-.13, .78]	.32 (.26) [-.19, .83]	.77 (.22)** [.33, 1.21]	.77 (.20)** [.38, 1.16]
<u>Home Context</u>						
Skills	-.21 (.23) [-.67, .24]	-.21 (.20) [-.60, .18]	.42 (.21)* [.01, .84]	.42 (.17)* [.09, .76]	.46 (.20)* [.06, .86]	.46 (.13)** [.20, .72]
Skills + Contact	.12 (.23) [-.34, .58]	.12 (.25) [-.54, .12]	.65 (.21)** [.23, 1.07]	.65 (.21)** [.23, 1.07]	.83 (.20)** [.43, 1.24]	.83 (.18)** [.48, 1.18]

Note. Regression models for treatment group effects before and after wild cluster bootstrapping (Cameron, Gelbach, & Miller, 2008), at pre-test, post-test, and 6-month follow-up. Ethnicity and gender were controlled for in the models. Referent group = control group. All confidence intervals are calculated as 95% confidence intervals. *B* = treatment effect slope, *SE* = standard error, CI = confidence interval.

⁺p < .10 *p < .05. **p < .01. ***p < .001.

Table 6a

Means and Standard Deviations for Baseline Equivalency Assessment for Exclusion Judgments and Justifications by Ethnicity and Treatment Group

		Judgments by Context									
		Undifferentiated Exclusion		Group-Based Exclusion		Parent Sanctioned Group-Based Exclusion		Peer Sanctioned Group-Based Exclusion			
		Peer <i>M(SD)</i>	Home <i>M(SD)</i>	Peer <i>M(SD)</i>	Home <i>M(SD)</i>	Peer <i>M(SD)</i>	Home <i>M(SD)</i>	Peer <i>M(SD)</i>	Home <i>M(SD)</i>		
<u>Ethnicity</u>		***	***	***	***	***	***	***	***		
Palestinian-Israeli/P-I		2.85(1.48)	2.20(1.38)	4.07(1.77)	3.12(1.71)	2.92(1.79)	2.18(1.38)	3.24(1.84)	2.53(1.55)		
Jewish-Israeli/J-I		3.99(1.55)	3.78(1.80)	5.33(1.46)	4.75(1.72)	4.48(1.87)	3.85(1.90)	4.75(1.77)	4.14(1.86)		
<u>Treatment Group</u>		-	-	-	-	-	-	-	-		
Control		3.62(1.56)	3.11(1.77)	5.01(1.60)	4.15(1.77)	3.91(1.97)	3.06(1.69)	4.19(1.98)	3.41(1.79)		
Skills		3.17(1.57)	2.71(1.68)	4.52(1.70)	3.71(1.86)	3.46(1.87)	2.87(1.79)	3.71(1.85)	3.10(1.78)		
Skills+Contact		3.53(1.69)	3.20(1.90)	4.63(1.87)	4.01(2.05)	3.79(2.11)	3.16(2.08)	4.15(2.02)	3.54(2.08)		
		<u>Justification by Context</u>									
		<u>Social-Conventional</u>		<u>Stereotype</u>		<u>Personal Choice</u>		<u>Moral</u>		<u>Excluder Empathy</u>	
		<u>Peer</u>	<u>Home</u>	<u>Peer</u>	<u>Home</u>	<u>Peer</u>	<u>Home</u>	<u>Peer</u>	<u>Home</u>	<u>Peer</u>	<u>Home</u>
<u>Ethnicity</u>		*	***	***	**	-	-	***	***	*	**
Palestinian-Israeli/P-I		0.57(0.43)	0.71(0.40)	0.53(0.50)	0.50(0.35)	0.40(0.49)	0.39(0.49)	0.17(0.30)	0.12(0.26)	0.15(0.36)	0.12(0.32)
Jewish-Israeli/J-I		0.45(0.41)	0.54(0.44)	0.27(0.44)	.35(0.48)	0.37(0.48)	0.34(0.48)	0.32(0.36)	0.28(0.36)	0.24(0.43)	0.23(0.42)
<u>Treatment Group</u>		-	-	-	-	-	-	-	-	-	*
Control		0.51(0.42)	0.60(0.43)	0.43(0.50)	0.39(0.49)	0.40(0.49)	0.40(0.49)	0.24(0.32)	0.22(0.34)	0.25(0.43)	0.24(0.43)
Skills		0.56(0.42)	0.69(0.43)	0.39(0.49)	0.45(0.50)	0.37(0.49)	0.33(0.47)	0.21(0.32)	0.17(0.30)	0.18(0.38)	0.19(0.39)
Skills+Contact		0.47(0.42)	0.57(0.43)	0.37(0.49)	0.43(0.50)	0.38(0.49)	0.37(0.49)	0.28(0.37)	0.22(0.34)	0.17(0.38)	0.10(0.30)

Note. Significance values for each follow-up comparisons is represented above each pair of means and standard deviations.

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; - = not significant.

Table 7a

Pairwise Comparisons of Means and Standard Error for Justifications by Gender, Ethnicity, and Scenario

	<u>Social-Conventional</u>	<u>Stereotype</u>	<u>Personal Choice</u>	<u>Moral</u>	<u>Excluder Empathy</u>
	<i>M(SE)</i>	<i>M(SE)</i>	<i>M(SE)</i>	<i>M(SE)</i>	<i>M(SE)</i>
Overall use of Justification	0.49(0.02) ^{a***}	0.29(0.02) ^b bc, bd*	0.38(0.02) ^c cb*; cd***	0.30(0.15) ^{bc}	0.20(0.16) ^d da, dc***; db*
<u>Gender</u>					
Female	0.43(0.03)	0.25(0.03)	0.37(0.03)	0.35(0.02)	0.23(0.02)
Male	0.55(0.03)	0.33(0.03)	0.38(0.03)	0.26(0.02)	0.18(0.02)
<u>Ethnicity</u>	***	***	*	***	**
Palestinian-Israeli	0.57(0.03)	0.37(0.03)	0.42(0.03)	0.22(0.02)	0.16(0.02)
Jewish-Israeli	0.42(0.03)	0.22(0.03)	0.33(0.03)	0.39(0.03)	0.25(0.02)
<u>Scenario</u>	***	-	-	***	-
Peer	0.44(0.02)	0.29(0.02)	0.36(0.02)	0.34(0.02)	0.20(0.02)
Family	0.54(0.02)	0.29(0.02)	0.38(0.02)	0.27(0.02)	0.20(0.02)

Note. Significance values for each pairwise comparison is represented above each pair of Means and Standard Errors; Different superscript letters for the “Overall use of Justification” row indicate significant mean differences across justifications; * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; - = not significant.